

Aus dem Physiologischen Institut
der Johannes Gutenberg - Universität Mainz

**INDIVIDUELL OPTIMALE UNTERSUCHUNGS-AUSWAHL
UND INDIVIDUELL OPTIMALE DIAGNOSTISCHE
AUSSAGE - EIN UMFASSENDES VERFAHREN ZUR
COMPUTER-GESTÜTZTEN DIFFERENTIALDIAGNOSE**

Inauguraldissertation
zur Erlangung der Würde des Doktors der Medizin
der Johannes Gutenberg-Universität in Mainz

dem Fachbereich Medizin
vorgelegt von
Clemens Friedrich H e B
aus Koblenz

MAINZ 1982



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41552619_0007

Aus dem Physiologischen Institut
der Johannes Gutenberg - Universität Mainz

**INDIVIDUELL OPTIMALE UNTERSUCHUNGS-AUSWAHL
UND INDIVIDUELL OPTIMALE DIAGNOSTISCHE
AUSSAGE - EIN UMFASSENDES VERFAHREN ZUR
COMPUTER-GESTÜTZTEN DIFFERENTIALDIAGNOSE**

Inauguraldissertation
zur Erlangung der Würde des Doktors der Medizin
der Johannes Gutenberg-Universität in Mainz

dem Fachbereich Medizin
vorgelegt von
Clemens Friedrich H e ß
aus Koblenz

MAINZ 1982

Aus dem Physiologischen Institut
des Johannes Gutenberg-Universitäts Mainz

Dekan: Professor Dr. K.H. Bäßler

1. Gutachter: Professor Dr. Dr. G. Thews

2. Gutachter: Professor Dr. H.J. v. Mengden

Tag der mündlichen Prüfung: 18. Juni 1982



Meiner Mutter

1. Die Welt ist ein großer Garten

2. Die Welt ist ein großer Garten

3. Die Welt ist ein großer Garten

4. Die Welt ist ein großer Garten

5. Die Welt ist ein großer Garten

6. Die Welt ist ein großer Garten

7. Die Welt ist ein großer Garten

8. Die Welt ist ein großer Garten

9. Die Welt ist ein großer Garten

10. Die Welt ist ein großer Garten

11. Die Welt ist ein großer Garten

12. Die Welt ist ein großer Garten

13. Die Welt ist ein großer Garten

14. Die Welt ist ein großer Garten

15. Die Welt ist ein großer Garten

16. Die Welt ist ein großer Garten

17. Die Welt ist ein großer Garten

18. Die Welt ist ein großer Garten

19. Die Welt ist ein großer Garten

INHALT

EINLEITUNG

KAPITEL I

ANALYSE DES DIAGNOSEVORGANGS	1
1.1. Ärztliche Diagnosefindung	1
1.2. Computer-gestützte Diagnose	3
1.2.1. Einleitung	3
1.2.2. Multivariate Diskriminanzanalyse	5
1.3. Grundzüge des neuen Verfahrens	12

KAPITEL II

MODIFIZIERTE DIAGNOSE	19
2.1. Einleitung	19
2.2. Methode	21
2.2.1. Modifizierte diagnostische Aussage	21
2.2.2. Fehldiagnoserate	21
2.2.3. Verlust an diagnostischer Schärfe	22
2.3. Ergebnisse	25
2.3.1. Reduktion der Fehldiagnoserate	26
2.3.2. Verlust an diagnostischer Schärfe	29
2.3.3. Der Modifikationsquotient	31
2.4. Ein zweites Beispiel: DD des Ikterus	35
2.5. Schlußfolgerungen	39

KAPITEL III

SEQUENTIELL INDIVIDUELLE UNTERSUCHUNGS-AUSWAHL	41
3.1. Einleitung	41
3.2. Methode	43
3.2.1. Theoretische Grundlagen	43
3.2.2. Multivariat Individuell Sequentielles Selektionsverfahren	44
3.3. Ergebnisse	51
3.3.1. Demonstration an Einzelfällen	51
3.3.2. Eindeutige diagnostische Aussage: Vergleich der Selektionsverfahren	61
3.3.3. Modifizierte diagnostische Aussage: Vergleich der Selektionsverfahren	65
3.4. Schlußfolgerungen	70

KAPITEL IV

WEITERE EIGENSCHAFTEN UND VARIATIONSMÖGLICHKEITEN DES NEUEN VERFAHRENS	73
4.1. Einfluß der Basisuntersuchungen	73
4.1.1. Einleitung	73
4.1.2. Eindeutige diagnostische Aussage	75
4.1.3. Modifizierte diagnostische Aussage	77
4.2. Wichtung von Krankheitsklassen	81
4.2.1. Einleitung	81
4.2.2. Theoretische Grundlagen	81
4.2.3. Ergebnisse	83

4.3. Andere Variationsmöglichkeiten	89
4.3.1. Ausschluß von Untersuchungen	89
4.3.2. Nicht ausreichende initiale differentialdiagnostische Aussage	90
4.3.3. Vorliegen mehrerer Erkrankungen	90
4.3.4. Berücksichtigung von Risiken und Kosten einer Untersuchung	91
4.3.5. Sequentielle Auswahl der besten Untersuchungskombinationen	93
4.4. Beziehungen des neuen Verfahrens zur ärztlichen Diagnosefindung	94
ZUSAMMENFASSUNG	98
Liste der Abkürzungen	101
Liste der Bezeichnungen	102
LITERATURVERZEICHNIS	103

EINLEITUNG

"Individuell optimale Untersuchungsauswahl und individuell optimale diagnostische Aussage" - dies sind zweifellos zwei sehr anspruchsvolle Forderungen an ein - noch dazu "umfassendes" - Verfahren zur Diagnosefindung (DF). Schließlich ist es das Ziel eines jeden Arztes, individuell für jeden Patienten die Untersuchungen (U) in geeigneter Art und Reihenfolge durchzuführen, dabei möglichst risikoarm und kostengünstig zu arbeiten und die erhobenen Resultate in eine ausreichend differenzierte diagnostische Aussage (DA) umzusetzen. "Optimal" kann dies natürlich nie sein und "optimal" ist auch die hier vorgestellte Methode sicherlich nicht. Was diesen Begriff aber in einem - provokativen - Titel rechtfertigt, ist die Tatsache, daß unser neues Verfahren sich einerseits weitgehend auf theoretische Optimalitätskriterien stützt und sich andererseits in den hier behandelten Beispielen allen vergleichbaren Methoden überlegen zeigt.

Allerdings beschränkt es sich auf die Differenzierung einer fest vorgegebenen initialen differentialdiagnostischen Aussage (DDA) mit Hilfe eines fest vorgegebenen maximalen "Pools" von U-en. Es erforscht, welche U-en hieraus in jedem individuellen Fall geeignet und nötig sind, um unter einer vorgegebenen Beschränkung der Fehldiagnoserate (FDR) eine - für den gegebenen "Pool" - möglichst weitgehend differenzierte DA zu treffen. In diesem Sinne gelten also alle Ergebnisse nur für diese klar umrissene Situation der Vorgabe der differentialdiagnostisch (DD) möglichen Erkrankungen und der maximal möglichen U-en.

Somit kann die neue Methode auch nur jeweils einen Teil eines gesamten Diagnosevorgangs erfassen, innerhalb dessen sie versucht, mit dem geringsten Aufwand an U-en eine möglichst weitgehende Differenzierung der initialen DDA bei gleichzeitig begrenzter FDR zu erreichen.

Da solche klar definierten Aufgabenstellungen der Behandlung mit Hilfe des Computers in besonderem Maße zugänglich sind (GROSS, 1966, DE DOMBAL, 1975), soll unser Verfahren zunächst als ein Computer-gestütztes Verfahren aufgefaßt werden, dessen Güte in Konkurrenz zu bekannten Computer-gestützten Methoden zu sehen ist und das daher im Verlauf der Arbeit immer wieder mit diesen verglichen wird. Dabei steht im Vordergrund der Vergleich mit der Methode der "Multivariaten Diskriminanzanalyse", die - von PIPBERGER in der automatischen Ekg-Analyse erstmalig benutzt - Ergebnisse der mathematischen Entscheidungstheorie nutzend, der Computer-gestützten Diagnose zum entscheidenden Durchbruch verholfen hat. (VON MENGDEN, 1977A). Sie ermöglichte als erste geringere FDR-en als bei der vergleichenden Klassifikation erfahrener Ärzte, zeigte jedoch als wesentliche Nachteile weder eine Flexibilität in der Auswahl der U-en noch in der Form der DA.

Diese beiden Eigenschaften der ärztlichen DF (äDF) führt unser Verfahren nun neu in die Computer-gestützte DF (CDF) ein, ohne jedoch die Vorteile der herkömmlichen "Multivariaten Diskriminanzanalyse" aufzugeben. Dabei zeigt es sich, daß es in der Lage ist, trotz Einsparung von U-en die FDR noch weiter zu reduzieren als diese.

Gleichzeitig stellt es aufgrund seiner großen Ähnlichkeit zur äDF eine Vergleichbarkeit der CDF mit dieser wieder her, so daß auch einige Eigenschaften des ärztlichen Diagnoseverhaltens analysiert werden können. So liefert unser Verfahren Hinweise darauf, welche ärztlichen Verhaltensweisen in welchen Diagnosesituationen ratsam sind. Da einer Übersetzung menschlicher Vorgehensweise in eine formalisierte Computer-kompatible Form naturgemäß enge Grenzen gesetzt sind, sind solche Schlußfolgerungen vorsichtig zu werten, gleichwohl enthalten sie aber einige aufschlußreiche Gesichtspunkte.

Die Art und die Vielzahl der neuen Aspekte unserer Methode macht es nötig, sie an einer übersichtlichen initialen DDA mit einem begrenzten "Pool" von U-en durchzuführen. Hierfür eignete sich besonders die DD von vier kongenitalen Herzvitien (Vorhofseptumdefekt, Pulmonalstenose, Fallotsche Tetralogie, Ventrikelseptumdefekt) mit Hilfe des Ekg, wobei ein "Pool" von neun Ekg-U-en (= "Ekg-Merkmale") benutzt wird.¹⁾ Die Güte des neuen Verfahrens wird getestet an den Ekg von 80 Kindern, deren tatsächliche Krankheit durch externe Methoden (z.B. Herzkatheter-U-en) gesichert wurde, vor unserer Klassifikation aber nicht bekannt war. So kann insbesondere bestimmt werden, welche FDR unsere Methode - im Vergleich zu anderen - zur Folge hat. Dies ist auch deshalb möglich,

1) Für die Überlassung der entsprechenden Daten danke ich Herrn Prof. Michaelis vom Institut für Medizinische Statistik und Dokumentation der Universität Mainz.

da schon von einigen Autoren (MICHAELIS, 1972, WATANABE, 1976) gezeigt worden ist, daß die Computer-gestützte-Ekg-DD kongenitaler Herzvitien niedrige FDR-en zu erbringen vermag, niedrigere etwa als mit einer Computer-gestützten Auswertung von Röntgenbildern (BROOKS, 1974).

Gemäß des Ansatzes unseres Verfahrens steht am Anfang dieser Arbeit eine kurze Analyse der äDF, wobei vier wünschenswerte Eigenschaften (I-IV) hervorgehoben werden: ein schneller (II) und stetiger (III) Diagnoseverlauf, der in eine ausreichend differenzierte (I) und vor allem zutreffende (IV) DA mündet. Dem gegenübergestellt wird unter den Methoden der CDF die der "Multivariaten Diskriminanzanalyse", die stets alle U-en eines vorgegebenen "Pools" benutzend i.a. genau eine der DD-vorgegebenen Erkrankungen auswählt, dabei Eigenschaft (IV) in hervorragender Weise genügt, in Bezug auf (I-III) aber keinerlei Variabilität erlaubt, so daß sie ein "Nicht-Sequentielles Selektionsverfahren" (NSV) mit stets eindeutiger DA (eDA) darstellt.

Variiert man jedoch für jeden Patienten den jeweils vorgegebenen U-Pool, etwa durch schrittweise Hinzunahme einer neuen U, kann man durch eine Folge von NSV-en ein sequentielles U-Verfahren (SV) entwickeln. Betrachtet man auf jeder Stufe dieses SV zusätzlich zu der eDA der "Multivariaten Diskriminanzanalyse" die jeweiligen (a posteriori) Wahrscheinlichkeiten jeder Krankheitsklasse, kann man auch die sich ergebende neue DDA formulieren. Da diese im Vergleich zur initialen DDA einige Erkrankungen

ausschließt, andere aber nicht zu unterscheiden vermag, wird sie als reservierte oder modifizierte DA (mDA) bezeichnet. Nun, durch die Konstruktion eines SV mit mDA, wird auch wieder eine Beurteilung der CDF in Form der Eigenschaften (I-IV) der äDF möglich und eine der äDF ähnliche Variabilität erreicht. Wie groß diese ist, zeigen (Kap. 1.3.) mehrere Ekg-Muster herzkranker Kinder; alle fünf Beispiele von Diagnoseverläufen erheben spezifische, teilweise gegensätzliche Forderungen an eine Selektion der U-en, eine Beendigung des SV und an die Klassifikationskriterien im Sinne einer mDA.

Die Durchführung dieser mDA, die in Kapitel 2 diskutiert wird, ist sehr einfach: Eine DD-vorgegebene Erkrankung wird, nach Vorliegen bestimmter U-Ergebnisse, als möglich angenommen, wenn ihre a posteriori Wahrscheinlichkeit eine vorgegebene Klassifikationskonstante (hier 10%) übersteigt. Die mDA besteht dann in der Auswahl einer Kombination der so bestimmten Erkrankungen - eine weitergehende Differenzierung bleibt weiteren U-en vorbehalten. Natürlich verbleibt auch bei dieser reservierten Diagnosehaltung ein Risiko der Fehldiagnose, jedoch kann es durch die Vorgabe der Klassifikationskonstanten von vorneherein beschränkt werden. Andererseits bedeutet der Verzicht auf weitergehende Differenzierung einen Verlust an diagnostischer Schärfe (VaDS), der ebenfalls quantifiziert werden muß. Dies geschieht durch die Konstruktion von zwei Größen, der Rate "echt mDA-en" (dies sind solche mDA, die mehr als eine Krankheit auswählen) und eines weiteren Maßes, des "Informationsgehalts" einer DA.

Die Durchführung dieser mDA bei der DD kongenitaler Herzvitien zeigt nun auf, was diese Klassifizierungsmethode im Vergleich zur eDA zu leisten imstande ist. Dazu werden verschiedene NSV-en bei unterschiedlicher Vorgabe des U-Pools betrachtet. Bei dem Vergleich der FDR-en von eDA und mDA fällt zunächst auf, daß beide ab einer bestimmten U-Anzahl sich kaum noch ändern, daß die FDR, die man bei Benutzung aller neun U-en erhält, schon mit fünf geeigneten U-en fast erreicht wird. Außerdem nimmt die Reduktion der FDR der eDA durch die Anwendung der mDA mit Zunahme der U-Anzahl monoton ab, was auch der praktischen Erfahrung entspricht: Je geringer die Zahl der benutzten U-en, umso sinnvoller ist eine reservierte Diagnosehaltung. Allerdings ist dann auch das Differenzierungsvermögen sehr gering und nimmt - mit steigender U-Anzahl und steigender FDR der mDA - zu. Sind diese beiden widerstrebenden Größen - FDR der mDA und V_{aDS} - auch getrennt voneinander gut zu quantifizieren, so ist es doch wichtig, sie in einem beide verbindenden Kriterium zu untersuchen. So gibt das hierzu konstruierte Maß, der Modifikationsquotient (MQ), an, wieviele Fälle man unnötigerweise mit einer mDA reserviert beurteilen muß, um eine Fehldiagnose der eDA zu korrigieren.

Bei der Benutzung eines Pools von 6 geeigneten Ekg-U-en ergibt sich etwa für die DD kongenitaler Herzvitien eine Reduktion der FDR der eDA durch die mDA von $15/80$ auf $4/80$ und eine Rate von $27/80$ echt-mDA-en, so daß ein MQ von $16/11$ erreicht wird. Dies unterstreicht die

Leistungsfähigkeit der mDA für das NSV: Sie reduziert die FDR der eDA auf fast ihren vierten Teil, für die Korrektur einer Fehldiagnose werden durchschnittlich ca 1,5 andere Fälle unnötigerweise reserviert beurteilt.

Allerdings wird der MQ alleine den diversen praktischen Situationen nicht gerecht, in einigen muß die Senkung der FDR absoluten Vorrang haben, in anderen eine weitgehende Differenzierung der initialen DDA. Zu steuern ist eine solche Schwerpunktsetzung über die Vorgabe der Klassifikationskonstanten, die eine Schranke für die FDR automatisch vorgeben. Unter Berücksichtigung dieser Begrenzung der FDR ermöglicht die mDA aber stets eine Aussage darüber, welches Differenzierungsvermögen mit dem vorgegebenen Pool an U-en zu realisieren ist.

Sind in Kapitel 2 bei der Diskussion der mDA anhand der NSV-en U-Auswahl und U-Anzahl (wenn auch in unterschiedlicher Weise) jeweils fest vorgeschrieben und für alle getesteten Ekgs stets jeweils gleich, so werden sie bei Anwendung der SV-en in Kapitel 3 variiert. Diese Variation besteht darin, für jeden individuellen Fall Schritt für Schritt, in Abhängigkeit der bereits erzielten U-Ergebnisse, die jeweils nächstbeste U auszuwählen und den Diagnosevorgang "zur rechten Zeit" zu beenden. Dies erfordert einerseits ein geeignetes Selektionsverfahren, andererseits eine Stopppregel, die die Durchführung aller für den betrachteten Fall relevanten U-en erlaubt, die der unnötigen aber unterläßt. Die Basis des hier konstruierten SV stellen neue Ergeb-

nisse der mathematischen Entscheidungstheorie dar (HESS, 1979A), die es gestatten, vor der Durchführung einer U ihren Nutzen nur aufgrund vorangegangener U-Resultate und der zugrundeliegenden statistischen Parameter abzuschätzen. Der Grundgedanke der eingeführten Stoppregel ist, die DF dann abzuberechnen, wenn in ihrem Verlauf selbst die beste U nicht in der Lage ist, eine weitere Differenzierung der DDA zu gewährleisten. Die Idee, die Stabilität der DF bei der CDF als Abbruchkriterium zu berücksichtigen, stellt ebenfalls ein neues und, wie sich zeigt, sehr nützliches Element des neuen Verfahrens dar.

Das SV wird in vier Formen vorgestellt; in seiner umfassendsten Fassung, dem "multivariat individuell sequentiellen Selektionsverfahren" (miSV) sucht es auf jeder Stufe der DF diejenige U aus, die in Kombination mit den bereits durchgeführten U-en ("multivariat") die größte Trennfähigkeit erwarten läßt, wobei dem individuellen Fall durch die Berücksichtigung der mit Hilfe der bereits bekannten U-Ergebnisse berechneten a posteriori Wahrscheinlichkeiten Rechnung getragen wird. Die übrigen diskutierten SV-en stellen Einschränkungen des miSV dar: Bei ihnen erfolgt die U-Auswahl nur durch Vergleich der Trennfähigkeit der U-en im einzelnen ("univariat") oder die auf jeder Stufe zuletzt berechneten a posteriori Wahrscheinlichkeiten und damit der aktuelle Entscheidungsstand bleiben unberücksichtigt ("durchschnittliches" SV), woraus jeweils ein geringerer Rechenaufwand bei der Durchführung dieser SV-en resultiert. Am Schluß der

DF steht bei allen SV-en wie bei dem NSV jedoch immer eine multivariate DA, die entweder als eDA oder als mDA die statistischen Abhängigkeiten der benutzten U-en mit einbezieht. Im Gegensatz zum NSV ist bei den SV-en die für jeden Patienten benötigte U-Anzahl verschieden, bei den "individuellen" SV-en zusätzlich auch die Reihenfolge der U-en.

Beim Testen der SV-en anhand von 80 Ekgs ergeben sich im Vergleich zum NSV teilweise erstaunliche Resultate. Bei Benutzung der eDA liegt der Vorteil der SV-en im wesentlichen in einer zum Teil erheblichen Ersparnis an U-en. Meist werden zum Erreichen der niedrigsten FDR der NSV ($11/80$ bei Benutzung aller 9 U-en) für das miSV 5 bis 6 U-en benötigt. Bei Anwendung der mDA kommt zu der Ersparnis an U-en noch eine weitere Reduktion der FDR bei einem nur geringen VaDS hinzu. Beginnt man etwa mit zwei ausgewählten Basis-U-en und führt das miSV dann schrittweise aus, bekommt man bei einer durchschnittlichen Anzahl von 5,8 U-en nur eine FDR der mDA von $3/80$ ($11/80$ bei eDA und 9 U-en). Um diese Korrektur von $8/80$ Fehldiagnosen zu erreichen, muß man gleichzeitig nur $12/80$ Ekgs unnötigerweise reserviert beurteilen.

Vergleicht man miSV und NSV bei jeweils gleicher durchschnittlicher U-Anzahl, fällt auf, daß das miSV besonders bei der Diskussion des Differenzierungsvermögen (VaDS) viel besser abschneidet, was sich in einem niedrigen MQ äußert. Neben dieser Tatsache, die die mDA als für eine schrittweise U-Auswahl besonders nützlich ausweist, ist erstaunlich, wie weit miSV und mDA gemeinsam die

FDR der "Multivariaten Diskriminanzanalyse" zu verringern vermag, obwohl (oder weil) nicht die Trennkraft aller zur Verfügung stehenden U-en ausgeschöpft wird. Die Betrachtung des miSV anhand der in Kap. 1.3. angeführten Beispiele möglicher Diagnoseverläufe zeigt nämlich, daß es oft gerade eine zu große U-Anzahl ist, die die DF in die falsche Richtung bringt und so zu einer größeren FDR führt - eine Erfahrung, die auch bei der äDF gemacht werden kann: eine größere Zahl von U-en ergibt häufig keine größere diagnostische Klarheit, nicht selten sogar eine größere Konfusion.

Wie die Resultate zeigen, ist das miSV gut in der Lage, dies zu verhindern. Es bietet eine rationelle U-Auswahl, indem es angibt, welche U-en aus einem vorgegebenen "Pool" tatsächlich für die DA in jedem individuellen Fall nötig sind und wie weit mit Hilfe dieses "Pools" die initiale DDA differenziert werden kann, so daß die Leistung des miSV für den betreffenden Teil des gesamten Diagnosevorgangs klar zu definieren und zu erkennen ist.

Ein weiterer Vorteil des miSV ist, daß seine FDR weitgehend unabhängig ist von der Auswahl der Basis-U-en, der U-en, mit denen man das SV beginnt. Dies liegt darin begründet, daß die darauf folgende optimale U-Auswahl etwaige Fehler zu Beginn der DF gut kompensieren kann (Kap.4.1.). Dazu kommt, daß die Anwendung einer mDA eine zusätzliche Wichtung schwerwiegender Erkrankungen, die bei der eDA zu deren Erkennung nötig ist, überflüssig

macht, da sie gemäß ihrer Konstruktion sehr sensibel für das Vorliegen einer Krankheit ist (Kap. 4.2.).

Werden diese beiden Gesichtspunkte auch an Ergebnissen der DD kongenitaler Herzvitien erläutert, so sind die weiteren Variationsmöglichkeiten des miSV nur theoretisch beschrieben (Kap. 4.3.) Danach ist es leicht, bei dem neuen Verfahren gewisse U-en nicht zur Beurteilung heranzuziehen, sei es, weil sie an einer Klinik nicht zur Verfügung stehen, sei es, weil ihre Ausführung für den Patienten zu riskant oder ihre Unbrauchbarkeit wegen technischer Mängel sofort nach ihrer Durchführung offenkundig ist. Außerdem machen miSV und mDA gemeinsam es teilweise möglich zu erkennen, ob die initiale DDA die tatsächliche Krankheit eines Patienten überhaupt erfaßt oder ob mehrere Erkrankungen gleichzeitig vorliegen. Dagegen bereitet die quantitative Wichtung von Risiken und Kosten der U-en in praxi Schwierigkeiten, ist prinzipiell aber ebenfalls in das miSV einzufügen.

Überhaupt ist überraschend, wie einfach das dem miSV entsprechende Computerprogramm strukturiert ist, obwohl es eine große Flexibilität in der Ausführung gewährleistet und strengen Optimalitätsforderungen der mathematischen Entscheidungstheorie gerecht wird. Dabei benötigt es vorab dieselben grundlegenden Berechnungen wie die "Multivariate Diskriminanzanalyse" bei PIPBERGER: eine zuverlässige multivariate Schätzung der den Krankheiten zugrunde liegenden statistischen Verteilungen (mit Mittelpunktsektoren und Kovarianzmatrizen). Wegen der Ein-

fachheit des dem miSV entsprechenden Computerprogramms verzichten wir in dieser Arbeit vollständig auf die Darstellung seiner technischen Details, sondern beschränken uns auf die Vorstellung des miSV in Form eines Flußdiagramms und auf die Erläuterung seiner theoretischen Grundlagen und seiner praktischen Auswirkungen.

Die neuen SV-en sind aber nicht nur fähig, die FDR-en der bekannten Computer-gestützten Diagnoseverfahren zu verringern, sondern sie erlauben auch, Rückschlüsse aus Ergebnissen der CDF auf die äDF zu ziehen (Kap. 4.4.). Hierfür eignet sich besonders die Diskussion des "univariat individuell" SV - schrittweise Auswahl der jeweils einzeln am besten trennenden U in Abhängigkeit des jeweiligen Falls und terminale DA unter Berücksichtigung aller erzielter Ergebnisse - , entspricht dies doch der ärztlichen Verhaltensweise weitgehend. So können Resultate der CDF Hinweise darauf geben, in welcher Reihenfolge der Arzt im Einzelfall die U-en ausführen sollte und welches Gewicht er den einzelnen Ergebnissen innerhalb seiner Gesamtbeurteilung beimessen sollte.

Wichtig ist schließlich, daß das Computersystem der SV-en für den Arzt dialogfähig ist; er kann an jeder entscheidenden Stelle in den Verlauf der DF eingreifen und neue Schwerpunkte setzen, wobei das Computerprogramm in der Zwischenzeit nach den vorgegebenen Optimalitätskriterien arbeitet. So kann der Arzt das Programm unterbrechen, eine nicht Computer-compatible U ausführen und mit einer danach geänderten DDA das SV fortführen. Damit

ist der Arzt in der Lage - und dies scheint bei der zunehmenden Bedeutung der Computertechnologie wesentlich - zu entscheiden, welche Teile eines Diagnoseverlaufs mit Computerunterstützung durchgeführt werden sollten, und zu erkennen, was das benutzte Verfahren dabei zu leisten imstande ist. So kann er der Maschine genau die Aufgaben zuweisen, bei deren Bewältigung sie ihm überlegen ist, und sich auf die Probleme konzentrieren, die menschliche Fähigkeiten unabdingbar benötigen.

Bemerkt sei noch, daß zur besseren Orientierung am Schluß der Arbeit eine Liste der verwendeten Bezeichnungen und Abkürzungen aufgeführt ist.

Mein besonderer Dank gilt Herrn Dr. K. Brodda, dessen zahlreiche Anregungen für den Fortgang der Arbeit außerordentlich hilfreich waren. Finanziell unterstützt wurde ich von Januar bis Juni 1981 durch ein vom Senat der Universität Mainz bewilligtes Stipendium.

Schließlich danke ich Herrn Prof. Dr. Dr. Thews herzlich dafür, daß er die Betreuung dieser Dissertation übernommen hat.

KAPITEL I

ANALYSE DES DIAGNOSEVORGANGS

1.1. Ärztliche Diagnosefindung

Die ärztliche Suche nach der Krankheit eines Patienten verläuft i.a. in mehreren Schritten:

- Nach Erhebung von Anamnese und körperlichem Untersuchungsbefund erfolgt eine erste differentialdiagnostische Aussage (DDA).
- Nach Ausführung gewisser Routineuntersuchungen (Labor, Röntgen, Ekg) wird die DDA geändert oder eingeschränkt.
- Für die neue DDA werden zusätzliche, zunächst leicht verfügbare Untersuchungen (U) angefordert, die die weitere Einschränkung der DDA am ehesten erwarten lassen.
- Nach jedem (bzw. jeder Folge von) U-Ergebnis(sen) wird die DDA überprüft, gegebenenfalls geändert.
- Die Diagnosefindung (DF) wird beendet, falls
 - (i) die initiale DDA in ausreichendem Maße zu einer endgültigen terminalen diagnostischen Aussage (DA) eingeschränkt werden konnte (Dies gilt z.B. dann, wenn eine weitere Differenzierung der Krankheiten keine Konsequenzen bezüglich Therapie oder Prognose hat.),
 - (ii) die weiteren U, die zur Verfügung stehen, keine weitere Differenzierung der DDA erwarten lassen,

oder falls

(iii) alle möglichen U ausgeführt wurden.

Nach Abbruch der DF stellen sich dem Arzt zu ihrer Beurteilung folgende Fragen:

(I) Ist die getroffene terminale diagnostische Aussage eindeutig oder bleiben nicht zu tolerierende Unklarheiten? (EINDEUTIGKEIT)

(II) War die DF in ihrem Verlauf stetig und gleichmäßig oder sprunghaft mit häufig wechselnder DDA?
(GLEICHMÄSSIGKEIT)

(III) War die Geschwindigkeit der DF schnell - was die Anzahl der U bzw. die benötigte Zeit betrifft - oder schleppend und langsam? (GESCHWINDIGKEIT)

(IV) Ist die Diagnose richtig oder falsch?

Während Frage (IV) erst spät nach Ende der DF - durch den Krankheitsverlauf oder durch Autopsy - oder auch gar nicht beantwortet werden kann, können die übrigen Fragen sofort nach der Durchführung der letzten U und nach Wahl der terminalen DA abschließend beurteilt werden. Dabei beinhaltet Punkt (I), welche Schärfe die DA besitzt und welchen Informationsgewinn die U-en eingebracht haben, (II) und (III), ob die U sinnvoll und rationell geplant und in der richtigen Reihenfolge durchgeführt wurden.

Bevor diese Probleme der ärztlichen DF (äDF) anhand von Beispielen erläutert werden (Kap. 1.3.), soll die Vorgehensweise bei Computer-gestützter DF (CDF) aufgezeigt werden.

1.2. Computer-gestützte Diagnose

1.2.1. Einleitung

In Anlehnung an die in Kap. 1.1. dargestellten ärztlichen Vorgehensweisen sind zahlreiche "human-like-reasoning"-Methoden auch für die Computer-gestützte DF (CDF) entwickelt worden (z.B. MOON, 1977), so die häufig angewandte "decision-tree"-Methode (GRÉMY, 1977, PÖPPL, 1977) oder die "Fuzzy set theory" von SMETS et al. (1976). Allen diesen Verfahren gemeinsam ist, daß nach ihrer Ausführung stets eine hohe Fehldiagnoserate (FDR) zu beobachten war, daß diese "artificial intelligence", wie GRÉMY (1977) bemerkte, mehr Analogie zu ärztlichen Verhaltensweisen als Genauigkeit in der DA beinhaltete. So war sie nur selten in der Lage, die FDR ärztlicher DF zu verringern.

Dies änderte sich, etwa bei der Differentialdiagnose von Herz- und Lungenerkrankungen mit Hilfe des Ekg, als Pipberger und Mitarbeiter die Anwendung der "Multivariaten Diskriminanzanalyse" begannen. Hiermit gelang es erstmals, die diagnostische Qualität der Computer-gestützten Ekg-Auswertung - gemessen an der Größe der FDR - so zu verbessern, daß sie sich selbst der erfahrener Kardiologen überlegen zeigte (VON MENGDEN, 1977 A). Der Erfolg dieser "zweiten Generation" von Ekg-Klassifizierungsverfahren veranlaßte PIPBERGER (1973) ebenso wie GRÉMY (1977) dazu, "human like reasoning"-Methoden für die Zukunft keine große Bedeutung zu prophezeien. Erkauft wurde dieser Fort-

schritt - die erhebliche Reduktion der FDR - dadurch, daß der Entscheidungsablauf für den Arzt undurchsichtig, ein Vergleich von ärztlicher und Computer-gestützter DF weitgehend unmöglich wurde (PIPBERGER, 1978, FOX, 1980)

In dieser Arbeit soll nun versucht werden, diesen beiden Anliegen, einer möglichst geringen FDR einerseits und einer weitreichenden Vergleichbarkeit von ärztlicher und Computer-gestützter DF andererseits gerecht zu werden. Dabei zeigt es sich, daß man auf den Nutzen der multivariaten Diskriminanzanalyse prinzipiell nicht verzichten kann. Da sie, wenn auch in veränderter Form, eine Grundlage unseres neuen Verfahrens darstellt, soll sie zunächst vorgestellt werden. Wie in der Einleitung bemerkt, beschränken wir uns dabei (wie Pipberger) auf eine festumrissene Situation: die Differenzierung einer initialen DDA - repräsentiert durch die Vorgabe einer bestimmten Zahl von Erkrankungen - mit Hilfe eines "Pools" von U, wobei die zu differenzierenden Krankheiten durch externe, mit dem gegebenen U-pool disjunkte Kriterien definiert sind.

Als Beispiel dient uns die Differentialdiagnose (DD) von vier kongenitalen Herzvitien (1. Vorhofseptumdefekt = ASD, 2. Pulmonalstenose = PS, 3. Fallotsche Tetralogie = FT, 4. Ventrikelseptumdefekt = VSD¹⁾) mit Hilfe von maximal neun Ekg-U (= Ekg-Merkmalen); das Vorliegen eines der Herzvitien ist definiert durch Herzkatheter-U-en oder Autopsy (MICHAELIS, 1972)

1) Im folgenden sind Krankheits-Nummern (j = 1, 2, 3, 4) und Krankheitsbezeichnungen (ASD, PS, FT, VSD) äquivalent.

1.2.2. Multivariate Diskriminanzanalyse

Trifft man zwischen N verschiedenen Krankheiten (hier $N = 4$) eine differentialdiagnostische Entscheidung, so steht am Ende häufig eine Aussage über die Wahrscheinlichkeit, mit der jede der in Frage kommenden Erkrankungen anzunehmen ist ¹⁾. Dazu wird in unserem Beispiel für jeden Patienten P das Ekg x registriert. Hieraus werden "geeignete" Größen für die gegebene DDA bestimmt. Bei unserer DDA sind dies die in Tab. 1.1. gezeigten Ekg-Merkmale (vgl. MICHAELIS, 1972).

Tab. 1.1.:

Nr.	Ekg-Merkmale
1	α - QRS ($\star$)
2	S I (mv)
3	S II (mv)
4	S III (mv)
5	"a" in Abl. I (mv)
6	"b" in Abl. I (mv)
7	RaVR (mv)
8	Max ($RV_1, R'V_1$) (mv)
9	SV ₁ (mv)
10	SV ₅ (mv)
11	SV ₆ (mv)

Ekg-Merkmale bei konventioneller Ekg-Registrierung
für die Computer-gestützte Differentialdiagnose kongenitaler Herzvitien (nach MICHAELIS, 1972).

1) Nach GROSS (1966) ist im Prinzip jede ärztliche Diagnose eine solche "Wahrscheinlichkeitsdiagnose".

Hat man etwa für einen Teil dieses vorgegebenen maximalen Pools von M (hier $M = 9$)¹⁾ Merkmalen (PM), die M_a aktuellen Merkmale AM (z.B. $AM = \{5, 6\}$), die entsprechenden Werte des Ekgs x berechnet, erhält man einen aktuellen Ekg-Vektor $x^{(AM)} = (x_1^{(AM)}, \dots, x_{M_a}^{(AM)})$ (z.B. $x^{\{5,6\}} = (x_5, x_6)$) aus M_a Elementen, mit dessen Hilfe diese Wahrscheinlichkeiten jeder Krankheitsklasse bestimmt werden können. Man benötigt dazu die relativen Häufigkeiten (a priori Wahrscheinlichkeiten p_j , $j = 1, \dots, N$) jeder Klasse und die (bedingten) Wahrscheinlichkeiten $f (x^{(AM)} | j)$, mit der die aktuellen Werte des Ekgs x , $x^{(AM)}$, für jede Krankheit (Nr. j) angenommen werden.

Nach der "Bayes-Formel" bekommt man die Wahrscheinlichkeit $q (j | x^{(AM)})$, mit der Krankheit Nr. j für das gemessene Muster $x^{(AM)}$ anzunehmen ist:

$$q (j | x^{(AM)}) = p_j / f (x^{(AM)} | j) / f (x^{(AM)}) \quad \text{für } j = 1, \dots, N \quad \text{Gl. (1-1)}$$

$$\text{Dabei ist } f (x^{(AM)}) = \sum_{j=1}^N p_j f (x^{(AM)} | j) \quad \text{Gl. (1-2)}$$

die totale Wahrscheinlichkeit, mit der die Werte $x^{(AM)} = (x_1^{(AM)}, \dots, x_{M_a}^{(AM)})$ in dieser Konstellation auftreten.

Die relativen Häufigkeiten p_j bestimmt man anhand eines "Trainingskollektivs" von Ekgs, für die die tatsächlich zugrundeliegende Diagnose durch externe Methoden (ohne Ekg-Benutzung, s.o.) gesichert wurde²⁾. Hiermit kann man

1) Die Merkmale mit den Nr. 7 und 8 wurden nicht berücksichtigt, da sie zu nicht zu erklärenden "konfusen" Diagnoseverläufen führten.

2) Dies ist z.B. von PIPBERGER (1973) für eine 8-Klassen-Ekg Differentialdiagnose getan worden.

auch - bei ausreichender Fallzahl - die Verteilungen $f(x^{(AM)} | j)$ bestimmen. Nimmt man für diese Funktionen

"multivariate Normalverteilungen"¹⁾ an, so kann man sie schreiben als

$$f(x^{(AM)} | j) = |\det(\Sigma_j^{(AM)})|^{-1/2} (2\pi)^{-M_a/2} \exp\left\{-1/2 (x^{(AM)} - \mu_j^{(AM)})^T \Sigma_j^{(AM)-1} (x^{(AM)} - \mu_j^{(AM)})\right\} \quad \text{Gl. (1-3)}$$

Dabei stellen $\mu_j^{(AM)}$ die Mittelpunktvektoren der Krankheitsklassen Nr. j (hier mit je M_a Elementen) dar, die Größen $\Sigma_j^{(AM)}$ die Kovarianzmatrizen (je $M_a \times M_a$ Elemente), die nicht nur die Varianzen der M_a Merkmale, sondern auch ihre statistischen Abhängigkeiten untereinander berücksichtigen²⁾.

Für jeden beliebigen Pool AM aktueller Merkmale lassen sich diese Größen $\mu_j^{(AM)}$ und $\Sigma_j^{(AM)}$ leicht bestimmen, hat man einmal die Größen $\mu_j = \mu_j^{(PM)}, \Sigma_j = \Sigma_j^{(PM)}$ für den maximalen Pool PM von Merkmalen geschätzt. Man streicht dazu lediglich in μ_j und Σ_j alle die Zeilen und Spalten, die den aktuell nicht betrachteten Merkmalen entsprechen. So wird für

$$AM = \{5, 6\} \quad \Sigma_j^{(AM)} = \begin{pmatrix} \Sigma_j(5,5) & \Sigma_j(5,6) \\ \Sigma_j(6,5) & \Sigma_j(6,6) \end{pmatrix}.$$

-
- 1) Den Einfluß nicht normalverteilter bedingter Dichten auf die multivariate Diskriminanzanalyse hat u.a. LACHENBRUCH (1977) untersucht.
 - 2) Nimmt man dagegen an, daß die durchzuführenden U-en statistisch unabhängig sind, führt die Anwendung der "Bayes-Formel" (1-1) zu ungünstigen Klassifikationsergebnissen (JACQUES, 1976).

Die Problematik der Schätzung der Größen p_j , μ_j , Σ_j anhand eines Trainingskollektivs ist von vielen Autoren (z.B. CORNFIELD, 1973) ausführlich diskutiert worden, so daß sie in dieser Arbeit - bis auf den Einfluß auf die Größen p_j (vgl. Kap. 4.2.) - unberücksichtigt bleibt.

Die damit mögliche Berechnung der a posteriori Wahrscheinlichkeit $q (j | x^{(AM)})$ für das aktuelle Ekg-Muster $x^{(AM)}$ reicht aber für die klinische Praxis häufig nicht aus. Therapiewahl und Prognoseaussagen erfordern meist eine klare Stellungnahme dazu, ob eine der vorgegebenen Krankheiten vorliegt oder auszuschließen ist. Die Wahl einer "randomisierten" Entscheidung in der Form der Angabe der $q (j | x^{(AM)})$ genügt dazu jedoch nicht.

Trifft man aber etwa mit Hilfe einer Diagnoseregeln δ_E für $x^{(AM)}$ die DA $\delta_E (x^{(AM)}) \in \{1, \dots, N\}$, ordnet man $x^{(AM)}$ genau eine der vorgegebenen Krankheiten der DDA zu (eindeutige DA = eDA). Die bei der Anwendung von δ_E zu erwartenden FDR bekommt man dann durch die folgenden Gleichungen:

$$FDR_{\delta_E} (i | j) = \int \{ \delta_E (x^{(AM)}) = i \} f (x^{(AM)} | j) dx \quad \text{Gl. (1-4)}$$

ist dabei die Rate derjenigen Ekgs, die tatsächlich Klasse Nr. j angehören, von δ_E aber in die mit Nr. i $\neq j$ eingeordnet worden.

$$FDR_{\delta_E} (j) = \sum_{i=1, i \neq j}^N FDR_E (i | j) \quad \text{Gl. (1-5)}$$

stellt die zu erwartende Rate derjenigen Ekgs dar, die tatsächlich der Klasse j angehören, von δ_E aber

irgendwie falsch zugeordnet werden. Will man etwa (vgl. Kap. 4.2.) die Krankheit Nr. j auf jeden Fall erkennen, sollte man δ_E so wählen, daß $FDR_{\delta_E}(j)$ möglichst klein wird. Meist geht dies auf Kosten der anderen $FDR_{\delta_E}(i)$ für $i \neq j$, die dann außergewöhnlich groß werden.

So wird man i.a. als entscheidendes Kriterium die durchschnittlich zu erwartende FDR,

$$FDR_{\delta_E} = \frac{N}{\sum_{j=1}^N} p_j FDR_{\delta_E}(j) \quad \text{Gl. (1-6)}$$

wählen.

Diejenige diagnostische Entscheidungsregel, die diese Größe minimal hält, ist die "Bayes-Entscheidung" (BE) δ_B .

Sie ist dadurch charakterisiert, daß sie einem Ekg-Muster $x^{(AM)}$ diejenige Krankheitsklasse i zuordnet, für die $q(i | x^{(AM)})$ (nach Gl. (1-1)) maximal ist, also

$$\delta_B(x) = i, \text{ falls } q(i | x^{(AM)}) \geq q(j | x^{(AM)}) \quad \text{Gl. (1-7)}$$

für alle $j = 1, \dots, N$.

Für die zugehörige FDR gilt dann:

$$FDR_B := FDR_{\delta_B} = \min_{\delta_E} FDR_{\delta_E} \quad \text{Gl. (1-8)}$$

Diese "Bayes-Entscheidung" δ_B wird bei der "multivariaten Diskriminanzanalyse" i.a. verwendet (MICHAELIS, 1972)

und ist, gemäß Gl. (1-8), für die niedrige FDR dieses Verfahrens im wesentlichen verantwortlich. Neben den schon erwähnten Einflüssen der Größen p_j , μ_j und Σ_j spielt für die Größe der FDR die Anzahl N der zur DDA stehenden Erkrankungen wie die Anzahl M_a der dem jeweiligen Pool von U angehörenden Merkmale eine Rolle. Ist der Einfluß der Klassenauswahl N, der in Kap. 2 von Bedeutung sein wird,

von HESS und BRODDA theoretisch (1979 A, B) und in seinen praktischen Auswirkungen (1979 B, 1980) ausführlich behandelt worden, so wird die Bedeutung der Ekg-Merkmalanzahl M_a in Kap. 3, bei der Auswahl optimaler U-Reihenfolgen, deutlich.

Insgesamt folgt die beschriebene DF in Form der eDA den von GROSS (1966) beschriebenen vier Schritten jeder ärztlichen DF:

- Speicherung typischer Daten für bestimmte Erkrankungen ("remember"): $p_j^{(AM)}, m_j^{(AM)}, \sum_j^{(AM)}$
- Eingabe der aktuellen Patientendaten: $x^{(AM)}$
- Vergleich der eingegebenen Daten mit den gespeicherten typischen Krankheitsdaten ("compute and compare"): $f (x^{(AM)} | j)$
- Angabe der Diagnose: $\delta_B (x^{(AM)}) = i$ und der Wahrscheinlichkeit, mit der sie zutrifft: $q (i | x^{(AM)})$ ("decide").

Die Regel δ_B zeichnet sich dabei durch fünf fundamentale Eigenschaften aus:

- (i) Sie ist ein streng nach statistischen Kriterien definiertes Verfahren.
- (ii) Sie benutzt *a l l e* U-en eines vorweg bestimmten aktuellen Pools von Ekg-Merkmalen.
- (iii) Sie berücksichtigt dabei ihre multidimensionalen Abhängigkeiten (Gl. (1-3)).
- (iv) Sie trifft immer eine eindeutige DA.
- (v) Sie stellt die Methode dar, die unter allen Verfahren mit den Eigenschaften (i) - (iv) die ge-

ringste durchschnittliche FDR erwarten läßt (Gl. (1-8)). Diese letzte Eigenschaft ist es auch, die der eDA in Form der BE ihre überragende praktische Bedeutung zukommen läßt.

Dagegen erlaubt die BE in ihrer multivariaten Form eine Variabilität der DF im Sinne der Punkte (I) - (IV) aus Kap. 1.1. nicht mehr. Da stets alle zur Verfügung stehenden U-en benutzt werden, ist die Geschwindigkeit der DF stets langsam, die Frage über die Gleichmäßigkeit des Diagnoseverlaufs nicht zu beantworten. Die Möglichkeit, bei schwierig zu klassifizierenden Fällen eine nicht eindeutige DA zu treffen, ist nicht gegeben.

1.3. Grundzüge des neuen Verfahrens

Das hier vorgestellte neue Diagnoseverfahren soll gerade die zum Schluß von Kapitel 1.2.2. erwähnten Nachteile der multivariaten Diskriminanzanalyse - Verzicht auf eine Variabilität in der U-Auswahl und in der Form der DA - beseitigen, ohne ihren Nutzen - die geringe FDR - aufzugeben. Die Ziele der neuen Methode sind somit,

- die FDR so gering wie möglich zu halten,
- dazu in einigen Fällen nicht eindeutige DA-en zuzulassen
- in jedem individuellen Fall (d.h. für jedes Ekg-Muster x) nur die U-en durchzuführen, die zu einer möglichst weitgehenden Differenzierung der initialen DDA notwendig sind.

Dazu ist es notwendig, in die CDF den Aspekt einer " modifizierten DA " und den einer " sequentiellen U-Auswahl " einzuführen. Der zweite Aspekt ist zu realisieren, indem man den jeweils betrachteten Pool aktueller Merkmale (AM) variiert und eine Folge von aktuellen Pools AM_j betrachtet, in der der erste (AM_0) fest vorgegeben wird und jeder folgende (AM_j) durch die Hinzunahme einer zusätzlichen U aus dem jeweils vorhergehenden (AM_{j-1}) entsteht.

In unserem Beispiel, der DD kongenitaler Herzvitien, beginnen wir mit den Ekg-Merkmalennummern 5 und 6 ($M_a=2$) und konstruieren die weiteren Pools durch schrittweises Hinzufügen der übrigen Merkmale - wobei der Reihenfolge der U-en an dieser Stelle der Arbeit noch keine Beachtung geschenkt werden soll.

Hat man nun ein Ekg x zur Klassifizierung vorliegen, kann man die (nach 1.2.2.) relevanten Größen der multivariaten Diskrimnanzanalyse für jeden der Pools der gegebenen Folge bestimmen. So bekommt man zunächst eine der Folge $AM_0, AM_1, \dots$ entsprechende Folge von a posteriori Wahrscheinlichkeiten $q (j | x^{(AM)})$ jeder Krankheitsklasse j (nach Gl. (1-1) und Gl. (1-3)) mit einer jeweils um 1 steigenden Merkmalsanzahl M_a .

Die Idee einer modifizierten DA (mDA) verwirklicht man, indem man - bei vorgegebenem U-Pool - statt einer eDA, die nur die Krankheit mit maximalem $q (j | x^{(AM)})$ auswählt, alle die Klassen in der DA berücksichtigt, deren a posteriori Wahrscheinlichkeit eine vorgegebene Größe (hier 0.1) übersteigt. Diese Krankheiten erkennt man in einem Entscheidungsvektor E am Wert 1 ($E(j)=1$), alle anderen am Wert 0. E repräsentiert somit die Reduktion der initialen DDA auf eine neue, durch die Ergebnisse der durchgeführten U-en eingeschränkte DDA.

Die so entstehende Folge von $q (j | x^{(AM)})$ und $E(j)^{(AM)}$ stellt dann den Verlauf der DF dar, den man erhält, wenn man schrittweise bestimmte zusätzliche U-Ergebnisse zur DA heranzieht, bis man schließlich alle möglichen U-en durchgeführt hat. Bei Ekg Nr. 36 etwa bekommt man die in Tab. 1.2. gezeigte Abfolge.

In diesem Beispiel verläuft die DF mit stets wachsender Wahrscheinlichkeit auf die richtige Diagnose ($j=2$: PS) zu. Auch die Wahrscheinlichkeit für $j=1$ (ASD) steigt

anfangs noch an, fällt dann aber gleichmäßig auf ver-schwindend geringe Werte ab. FT und VSD bleiben schon früh durch E unberücksichtigt und bekommen auch im wei-teren Verlauf der DF kein entscheidendes Gewicht mehr.

Führte hier nur eine ausreichend große Zahl von U-en zum Ausschluß aller nicht zutreffenden Krankheiten, so ist in anderen Fällen nur eine geringe Merkmalszahl hier-für notwendig. Bei Ekg Nr. 20 etwa ist diese Situation schon im ersten Schritt, nach zwei U-en erreicht (Tab. 1.3.), so daß nur 7/9 der maximal möglichen U-Anzahl benötigt wird.

Eine geringere Anzahl von Merkmalen verringert aber nicht nur den Aufwand der DF, sondern kann auch, wie Ekg Nr. 31 (Tab. 1.4.) zeigt, Fehldiagnosen verhindern. Hier scheint die DF der richtigen Entscheidung (PS) zunächst sehr schnell zuzustreben, verändert mit der zusätzlichen Verwendung von Nr. 4 aber plötzlich die Richtung und führt schließlich zur falschen DA.

Führte in den erläuterten drei Beispielen eine geeignete Auswahl der U-en alleine - auch bei Anwendung der eDA (Entscheidung für die Krankheit mit maximalem $q(j|x^{(AM)})$) und ohne die Modifizierung durch E - zur richtigen Diagnose, so zeigt das Beispiel des Ekgs Nr. 1, daß dies nicht immer ausreicht (Tab. 1.5.).

Die tatsächlich diesem Ekg zugrunde liegende Erkrankung wird durch die eDA an keiner Stelle der DF erkannt; nur eine modifizierte DA (mDA) vermöge E berücksichtigt sie bei geeigneter U-Auswahl. Die Benutzung aller neun Merk-

Tab. 1.2.

AM	M _a	neue U:Nr	q (j x ^(AM)) für j =				E(j) für j =			
			1	2	3	4	1	2	3	4
AM ₁	2	5,6	0.2499	0.3680	0.2485	0.1335	1	1	1	1
AM ₂	3	2	0.3021	0.5174	0.0418	0.1386	1	1	0	1
AM ₃	4	10	0.4362	0.5186	0.0207	0.0245	1	1	0	0
AM ₄	5	1	0.3663	0.5848	0.0300	0.0189	1	1	0	0
AM ₅	6	3	0.1808	0.8065	0.0000	0.0128	1	1	0	0
AM ₆	7	9	0.1245	0.8636	0.0000	0.0119	1	1	0	0
AM ₇	8	11	0.0431	0.9505	0.0000	0.0064	0	1	0	0
AM ₈	9	4	0.0197	0.9739	0.0000	0.0063	0	1	0	0

Verlauf der DF für Ekg-Nr. 36: Änderung der a posteriori Wahrscheinlichkeiten $q (j | x^{(AM)})$ und des Entscheidungsvektors E bei Variation des aktuellen Merkmal-Pools AM mit U-Anzahl M_a. Im Gegensatz zur eDA, bei der nur die Krankheit j mit maximalem $q (j | x^{(AM)})$ ausgewählt wird, berücksichtigt E (durch E(j)=1) alle die Klassen j, für die $q (j | x^{(AM)}) \geq 0.1$ gilt.

Beginnt man mit der Untersuchung der Ekg-Merkmale Nr.5 und Nr.6 und vergrößert den U-Pool schrittweise um das jeweils bezeichnete Merkmal, wird hier die DDA gleichmäßig bis auf die richtige Krankheitsklasse 2 (PS) reduziert.

Tab. 1.3.

AM	M _a	neue U:Nr	q (j x ^(AM)) für j =				E(j) für j =			
			1	2	3	4	1	2	3	4
AM ₁	2	5,6	0.0172	0.9014	0.0696	0.0119	0	1	0	0
AM ₂	3	2	0.0005	0.9995	0.0000	0.0000	0	1	0	0
AM ₃	4	3	0.0001	0.9999	0.0000	0.0000	0	1	0	0
AM ₄	5	9	0.0001	0.9999	0.0000	0.0000	0	1	0	0
AM ₅	6	11	0.0000	1.0000	0.0000	0.0000	0	1	0	0
AM ₆	7	4	0.0000	1.0000	0.0000	0.0000	0	1	0	0
AM ₇	8	10	0.0000	1.0000	0.0000	0.0000	0	1	0	0
AM ₈	9	1	0.0000	1.0000	0.0000	0.0000	0	1	0	0

Verlauf der DF für Ekg-Nr. 20 (Bezeichnungen wie Tab. 1.2.): Schon die Benutzung der Basismerkmale Nr.5 und Nr.6 ermöglicht den Ausschluß aller nicht zutreffenden Krankheiten $j \neq 2$. Dazu werden nur 2/9 der zur Verfügung stehenden U-en benötigt.

Tab. 1.4.

AM	M _a	neue		q (j x ^(AM)) für j =				E(j) für j =			
		U:Nr		1	2	3	4	1	2	3	4
AM ₁	2	5,6		0.1167	0.8027	0.0677	0.0129	1	1	0	0
AM ₂	3	2		0.0722	0.9233	0.0001	0.0043	0	1	0	0
AM ₃	4	3		0.0100	0.9883	0.0000	0.0017	0	1	0	0
AM ₄	5	11		0.0083	0.9909	0.0000	0.0007	0	1	0	0
AM ₅	6	9		0.0049	0.9947	0.0000	0.0004	0	1	0	0
AM ₆	7	1		0.0272	0.9722	0.0000	0.0006	0	1	0	0
AM ₇	8	4		0.0001	0.0026	0.9973	0.0000	0	0	1	0
AM ₈	9	10		0.0004	0.0101	0.9895	0.0000	0	0	1	0

Verlauf der DF für Ekg-Nr. 31 (Bezeichnungen wie Tab. 1.2.):

Nach Hinzunahme von Merkmal Nr.4 wird die zugrundeliegende Krankheit 2 nicht mehr berücksichtigt. Eine geringere U-Anzahl als 7 würde hier eine Fehldiagnose verhindern.

Tab. 1.5.

AM	M _a	neue		q (j x ^(AM)) für j =				E(j) für j =			
		U:Nr		1	2	3	4	1	2	3	4
AM ₁	2	5,6		0.3327	0.0095	0.0704	0.5874	1	0	0	1
AM ₂	3	10		0.3609	0.0135	0.0610	0.5645	1	0	0	1
AM ₃	4	1		0.0613	0.0012	0.0000	0.9375	0	0	0	1
AM ₄	5	4		0.0651	0.0000	0.0000	0.9349	0	0	0	1
AM ₅	6	11		0.1330	0.0000	0.0000	0.8669	1	0	0	1
AM ₆	7	2		0.1039	0.0000	0.0000	0.8971	1	0	0	1
AM ₇	8	9		0.0533	0.0001	0.0000	0.9467	0	0	0	1
AM ₈	9	3		0.0885	0.0000	0.0000	0.9114	0	0	0	1

Verlauf der DF für Ekg-Nr. 1 (Bezeichnungen wie Tab. 1.2.):

Die richtige Diagnose (j=1: ASD) kann nur mit Hilfe des Entscheidungsvektors E erfaßt werden, bei Anwendung der eDA ist eine Fehldiagnose bei jeder Wahl von AM unumgänglich.

Tab. 1.6.

AM	M _a	neue		q (j x ^(AM)) für j =				E(j) für j =			
		U:Nr		1	2	3	4	1	2	3	4
AM ₁	2	5,6		0.2412	0.0278	0.0398	0.6911	1	0	0	1
AM ₂	3	10		0.6411	0.0221	0.1065	0.2303	1	0	1	1
AM ₃	4	1		0.9038	0.0194	0.0058	0.1710	1	0	0	1
AM ₄	5	3		0.8039	0.0275	0.0000	0.1686	1	0	0	1
AM ₅	6	11		0.8287	0.0181	0.0000	0.0932	1	0	0	0
AM ₆	7	4		0.9114	0.0267	0.0000	0.0619	1	0	0	0
AM ₇	8	9		0.9047	0.0000	0.0000	0.0353	1	0	0	0
AM ₈	9	2		0.7311	0.0000	0.0000	0.2689	1	0	0	1

Verlauf der DF für Ekg-Nr. 2 (Bezeichnungen wie Tab. 1.2.):

Die Modifizierung der DA durch E fügt für M_a=4,5,9 der richtigen, durch die eDA klar zu erkennenden Klasse 1 (ASD) lediglich die Klasse 4 als mögliche Diagnose hinzu und führt zu einer unnötigerweise reservierten DA.

male führt allerdings auch bei Anwendung einer mDA nicht zum Erfolg.

Insgesamt zeigen die diskutierten Ekg-Beispiele, daß eine große U-Anzahl, obwohl in einigen Fällen nötig, häufig nicht nur überflüssig, sondern manchmal sogar für die Richtigkeit der DA schädlich sein kann. Es kommt bei der Konstruktion eines sequentiellen U-Verfahrens (SV) also zunächst darauf an, die DF an einer Stelle abzubrechen, an der die richtige Entscheidung möglich ist. Außerdem muß die schrittweise Auswahl der U-en so erfolgen, daß die DF einen möglichst schnellen und gleichmäßigen Verlauf nimmt, neigen doch - wie Ekg Nr. 1 andeutet (Tab. 1.5.) - wechselhafte Verläufe in besonderem Maße zur Entwicklung der DF in die falsche diagnostische Richtung.

Außerdem weist dieses Beispiel darauf hin, daß auch eine Modifizierung der eDA durch den Verzicht auf die Unterscheidung gewisser Krankheiten helfen kann, Fehldiagnosen zu vermeiden.

Somit erkennt man, daß eine größere Variabilität der CDF - im Sinne der Eigenschaften I - III (Kap. 1.1.) der äDF (Gleichmäßigkeit und Geschwindigkeit der DF, Verzicht auf die Eindeutigkeit der DA) - bei der Klassifizierung von U-Ergebnissen nützlich sein kann.

Daß dies bei geeigneter Konstruktion sequentieller U-Verfahren tatsächlich so ist, wird in Kap. 3 ausführlich erläutert. Bevor in Kap. 2 die Modifizierung der DA diskutiert wird, soll ein letztes Beispiel demonstrieren, daß ein solches - nicht eindeutiges - Diagnoseverhalten

aber auch Nachteile haben kann (Ekg Nr. 2, Tab. 1.6.).
Bei Benutzung von 4, 5 oder 9 U-en fügt die mDA hier der richtigen eDA (ASD) nur VSD als zusätzlich zu berücksichtigende Erkrankung hinzu, führt also zu einem unnötigerweise reservierten Diagnoseverhalten. Wie sich diese Vorteile (Reduzierung der FDR) und Nachteile (Verlust an diagnostischer Schärfe) der mDA bei der Behandlung von größeren Kollektiven (80 Ekgs) verhalten, wird in Kap. 2 diskutiert.

KAPITEL II

MODIFIZIERTE DIAGNOSE

2.1. Einleitung

Wie schon am Schluß von Kap. 1 erwähnt wurde, ist eine modifizierte DA (mDA), die mehrere der vorgegebenen Krankheiten zur endgültigen DA berücksichtigt, häufig in der Lage, Fehldiagnosen zu vermeiden, was allerdings stets mit einem Verlust an diagnostischer Schärfe erkaufte wird. Eine solche vorsichtige, reservierte Haltung findet man oft bei der ärztlichen DF; auch der Arzt ist nicht selten bemüht, eine eindeutige DA zu vermeiden, um sich nicht auf eine nicht zutreffende Krankheit festzulegen. Wie bei der CDF kann eine solche Einstellung hilfreich sein, in anderen Fällen ist sie aber unnötig und überflüssig. So stellen STERN et al. (1974), als sie die FDR-en von Computer und Ärzten bei der Differentialdiagnose des Ikterus verglichen, fest, daß die Anwendung einer mDA - hier in Form der Angabe der ersten drei in Frage kommenden Krankheiten - wohl die ärztliche DA erheblich verbesserte, nicht aber die FDR der CDF verringerte.

Da bei der äDF eine reservierte DA i.a. intuitiv getroffen wird, ist es schwierig, ihre Vor- und Nachteile zu beschreiben. Daher ist es auch nicht möglich, die Umstände anzugeben, unter denen eine mDA angewandt, oder die, unter denen sie besser unterlassen werden sollte. Hat man den Diagnoseverlauf jedoch - wie in Kap. 1.3. - in Form der konsekutiven DDA, mit den entsprechenden Wahrschein-

keiten $q(j|x^{(AM)})$, vorliegen, kann man sowohl die Kriterien für eine mDA genau festlegen, als auch ihre Vor- und Nachteile gegenüber der eDA gut erkennen. Hieraus wiederum können Schlüsse daraufhin gezogen werden, unter welchen Bedingungen eine mDA nutzbringend ist. Da diese Umstände, wie sich zeigen wird, leicht auf die ÄDF übertragen werden können, kann man auch für ärztlich ausgewählte U-Abläufe den Wert eines reservierten Diagnoseverhaltens beurteilen.

2.2. Methode

2.2.1. Modifizierte DA

Hat man sich mit Anwendung der eDA in Kap. 1.2.2. immer für eine Krankheitsklasse entscheiden müssen, so wird nun vorgesehen, die Unterscheidung zwischen bestimmten Erkrankungen in einzelnen kritischen Fällen offen zu lassen. Dies geschieht in der Art, daß man mit der DA alle die Krankheiten (K) angibt, für die eine ausreichend große Wahrscheinlichkeit (WK) $q (j | x^{(AM)})$ besteht (vgl. Kap. 1.3.). Ist J eine beliebige Teilmenge der Menge I aller N zur DD stehenden Krankheiten, so entscheidet man sich unter Kenntnis von allen M_a Merkmalen (eines vorgegebenen Pools AM) des Ekgs x, $x^{(AM)}$ für alle j aus J ($\delta_M (x^{(AM)}) = J$) falls gilt:

$$q (j | x) \geq c_j \quad \text{für alle } j \in J \quad \text{und}$$

$$q (j | x) < c_j \quad \text{für alle } j \notin J \quad \text{Gl. (2-1).}$$

Dabei sind die c_j (hier: $c_j = 0.1$ für $j = 1, \dots, N$) vorgegebene Größen, die die FDR der mDA δ_M von vorneher- ein beschränken:

2.2.2. FDR der mDA

Danach läßt sich die $FDR_M (J | k)$, sich für alle Krankheiten $j \in J$ und nicht für die tatsächliche richtige $k \notin J$ zu entscheiden, nach oben begrenzen:

$$P_k FDR_M (J | k) = \int_{\{ \delta_M (x^{(AM)}) = J \}} p_k f_k (x^{(AM)}) dx^{(AM)}$$

$$\leq c_k P \{ \delta_M (x) = J \}, \quad \text{Gl. (2-2)}$$

wobei $P \{ \delta_M (x) = J \}$ die WK ist, sich überhaupt für diese Kombination J von K-en aus I zu entscheiden.

Ähnliches bekommt man für die Rate $FDR_M (k)$, Ekgs, die tatsächlich der K k angehören, überhaupt falsch zuzuordnen,

$$p_k FDR_M (k) = \sum_{J: k \notin J} p_k FDR (J|k) \quad Gl. (2-3)$$

$$\leq \sum_{J: k \notin J} c_k p_k P \{ \delta_M (x^{(AM)}) = J \} \leq c_k p_k$$

Die durchschnittliche FDR der mDA kann man dann abschätzen durch

$$FDR_M = \sum_{k=1}^N p_k FDR_M (k) \leq \bar{c} := \sum_{k=1}^N c_k \quad Gl. (2-4).$$

Diese und andere theoretischen Eigenschaften der mDA sind von HESS (1979 A) auf der Grundlage von Untersuchungen von QUESENBERRY und GESSAMAN (1968) entwickelt worden. Sie setzen uns in die Lage, durch Vorgabe kleiner c_k die FDR_M sehr klein zu halten. Erkauft wird dieser Effekt aber damit, daß für viele Ekgs eine Kombination von K-en angegeben wird, im Extremfall sogar eine, die alle vorgegebenen K-en enthält und keine ausschließt.

2.2.3. Verlust an diagnostischer Schärfe (VaDS)

Die Quantifizierung dieses unerwünschten Effektes kann auf zweierlei Weise vorgenommen werden:

- (1) Man bestimmt beim Testen der mDA anhand eines Testkollektivs von n Probanden die Anzahl n_M derjenigen

Ekg's, für die eine echt mDA vorgenommen wird. (Echt mDA-en sind solche mDA-en, für die mindestens zwei K-en als mögliche Diagnose ausgewählt werden.) Als Kriterium für den VaDS bekommt man dann den Anteil $Q_n = n_M / n$ der echt mDA an alle DA insgesamt.

Der Nachteil dieses Kriteriums ist, daß es nicht berücksichtigt, wie unscharf die mDA tatsächlich ist. Es unterscheidet z.B. nicht zwischen mDA-en, die zwei K-en als möglich ansehen und solchen, die keine K ausschließen können.

- (2) Dieser Nachteil wird behoben durch die Einführung eines Informationsmaßes für die mDA. Hierfür bietet sich aus verschiedenen theoretischen Erwägungen (HESS, 1979 A) das "Shannonsche Informationsmaß" an:

Entscheidet man sich bei gegebenen $x^{(AM)}$ für alle K-en $j \in J$ ($\delta_M (x^{(AM)}) = J$), so erhält man die diagnostische Schärfe (DS) dieser mDA durch

$$\text{INF}_M (\delta_M (x^{(AM)})) = \text{INF}_M (J) = - \left(\sum_{j \in J} p_j \right) \ln \left(\sum_{j \in J} p_j \right) - \sum_{k \notin J} p_k \ln p_k \quad \text{Gl (2-5)}$$

Dieser Wert ist umso größer, je weniger K-en die Menge J enthält und weniger sich die Häufigkeiten p_j unterscheiden: $\text{INF}_M (\delta_M (x^{(AM)})) \leq \text{TINF}_E \leq \ln N$, wobei $\text{TINF}_E = \text{INF}_E (\delta_E (x^{(AM)})) = \sum_{j=1}^N p_j \ln p_j$ die DS jeder eDA, unabhängig von dem zu klassifizierenden Ekg-Muster x darstellt, da stets alle K-en von einer eDA differenziert werden ($\ln N$ ergibt sich als die DS bei gleichen $p_j = 1/N$).

Die totale (durchschnittliche) DS der mDA bei Untersuchung von n Ekgs $x_{(1)}^{(AM)}, \dots, x_{(n)}^{(AM)}$ ergibt sich demgegenüber als das Mittel der einzelnen (individuellen) DS-en:

$$TINF_M = \frac{1}{n} \sum_{i=1}^n INF (\delta_M (x_{(i)}^{(AM)})) \quad \text{Gl. (2-6).}$$

Als zweites Kriterium für den VaDS durch die mDA gegenüber der eDA erhält man dann die Verringerung des Informationsgehalts der eDA, $\Delta INF = TINF_E - TINF_M$.

Die praktische Anwendung beider Kriterien wird zeigen (Kap. 2.3.), daß die Beurteilung des VaDS nach (1) und (2) i.a. übereinstimmt und nur in wenigen Fällen unterschiedlich ist. Vergleicht man die Reduktion der FDR der eDA durch die mDA mit dem gleichzeitigen VaDS, so ist interessant, daß der relative Verlust an DS (durch die mDA) immer kleiner ist als die relative Reduktion der FDR (HESS, 1979 A):

$$\frac{\Delta INF}{TINF_E} \leq \frac{FDR_B - FDR_M}{FDR_B} \quad \text{Gl. (2-7)}$$

Nach dieser Formel ist jede nach Gl. (2-1) konstruierte mDA stets die bessere Alternative zu j e d e r eDA - was bei der praktischen Anwendung jedoch differenzierter betrachtet werden muß (Kap. 2.3. und Kap. 3.3.). Dabei stellt sich heraus, daß bei gewissen Diagnoseabläufen dieser Gewinn nur gering, bei anderen aber beträchtlich ist.

2.3. ERGEBNISSE der mDA

In diesem Paragraphen soll nun gezeigt werden, wie sich die Anwendung der mDA bei der Computer-gestützten DD kongenitaler Herzvitien verhält. Insbesondere soll - im Vergleich von mDA und eDA - geklärt werden,

- wieviele durch die eDA falsch diagnostizierte Ekgs durch die mDA korrigiert werden können (vgl. Bsp. Ekg-Nr. 1 in 1.3.) und
- wieviele durch die eDA richtig diagnostizierte Ekgs unnötigerweise durch die mDA echt modifiziert beurteilt werden (vgl. Bsp. Ekg-Nr.2 in 1.3.).

Dabei stehen uns als Kriterien

- die FDR-en beider Verfahren, FDR_B (Gl. (1-8)) und FDR_M (Gl. (2-4)) und
- der Anteil $Q_n = n_M / n$ echt modifizierter DA bzw. der $VaDF \Delta INF$ zur Verfügung.

Die nach Gl. (2-1) für die mDA notwendige Vorgabe der konstanten c_j sei hier mit $c_j = 0.1$ für alle K-en $j \in I$ getätigt¹⁾.

Vor- und Nachteile der mDA gegenüber der eDA sollen zunächst einzeln, später in einem beide verbindenden Maß, für die nicht sequentiellen Selektionsverfahren (NSV) diskutiert werden. Bei NSV werden die Ergebnisse aller U-en eines jeweils vorgegebenen aktuellen Pools g l e i c h z e i t i g zur DA herangezogen. Wir beobachten nun die Veränderung der FDR der eDA durch die mDA und den sich ergebenden VaDS in Abhängigkeit

1) Wie sich eine Variation dieser Klassifikationskonstanten auf die erwähnten Kriterien auswirkt und in welchem Zusammenhang sie zu diskutieren ist, ist u.a. Inhalt von Kap. 4.2..

von der jeweils benutzten Merkmalsanzahl M_a . Die FDR-en werden dabei bestimmt durch den Vergleich der für ein Kollektiv von 80 Ekgs jeweils getätigten DA mit den tatsächlich (durch externe Kriterien definierten) zugrundeliegenden Erkrankungen, der VaDS wird berechnet anhand der jeweils gemachten mDA (nach 2.2.3. (1), (2)).

Für jede feste Merkmalsanzahl M_a untersuchen wir diese Größen in erster Linie für die jeweils "beste" Merkmalsauswahl. Dieser für jedes M_a "beste" Pool von U - es gibt genau (M_a) m ö g - l i c h e Pools mit M_a U - wird bestimmt, indem man das bei schrittweiser Vergrößerung des Pools jeweils "beste" Merkmal nach dem in Kap. 3.3.2. beschriebenen "multivariat durchschnittlichen" Auswahlkriterium $CRIT_{MD}$ bestimmt, - wie in den Erörterungen in Kap. 4.1. gezeigt wird - stellt es auch wirklich das für jede Merkmalsanzahl beste NSV dar. So bekommt man den für M_a "besten" Pool von U durch die Merkmalskombination, die aus den ersten M_a U der Folge der Merkmals-Nr. 6, 2, 10, 1, 3, 5, 4, 9, 11 besteht. Die übrigen - nicht auf der durchgezeichneten Kurve der Diagramme 2.1. - 2.3. liegenden Punkte repräsentieren die Resultate anderer, beliebig ausgewählter Pools mit entsprechendem M_a .

2.3.1. Reduktion der FDR durch die mDA

Der Unterschied zwischen eDA und mDA in der FDR wird durch Diagramm 2.1 aufgezeigt. Dabei fällt auf, daß die FDR der eDA (FDR_B) mit wachsender Zahl an U-en (fast) stetig sinkt und von 48 % bis auf 13,75 % abfällt, im vorletzten Schritt sogar bei 12,5 % liegt. Dagegen wächst die FDR der mDA (FDR_M)

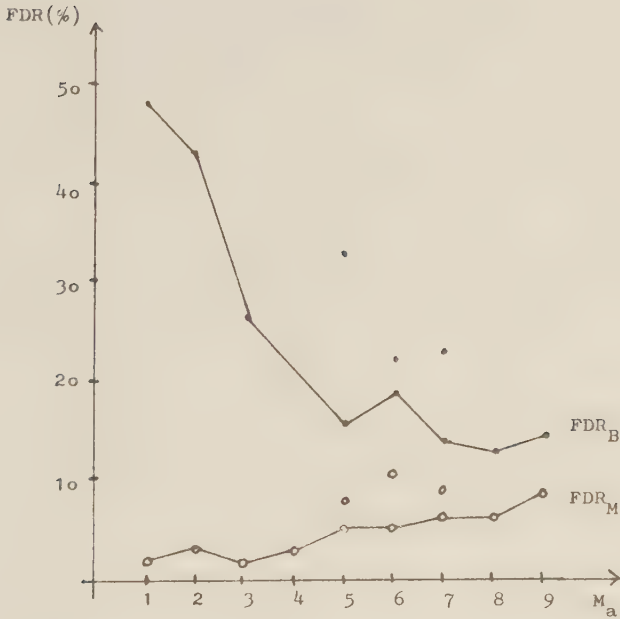


Diagramm 2.1.

Vergleich der FDR-en von eDA ($FDR_B, \bullet$) und mDA ($FDR_M, \circ$) für die jeweils "beste" U-Auswahl (durchgezeichnete Kurven) und einige andere Merkmalskombinationen . (Punkte) bei Variation der U-Anzahl M_a und Anwendung des NSV.

Während die Kurve für FDR_B monoton fällt (mit Ausnahme von $FDR_B(6)$), nimmt die für FDR_M (mit Ausnahme von $FDR_M(3)$) monoton zu, so daß die FDR der eDA durch die mDA mit steigender U-Anzahl immer weniger reduziert wird. Ab $M_a=5$ ändern sich FDR_B und FDR_M nur noch wenig.

mit wachsender Merkmalsanzahl langsam von 1,25 % auf 8,75 % an.¹⁾ Die Reduktion der FDR durch die mDA ist bei Benutzung nur einer U mit 43,75 % erheblich (die relative Reduktion $(FDR_B - FDR_M) / FDR_B$ nach Gl. (2-7) mit einem Wert von 97 % ebenfalls). Diese Reduktion fällt schließlich ab bis auf eine Größe von 5 % , wobei 36 % der FDR_B eingespart werden.

Bemerkenswert ist, daß ab einer Zahl von fünf U-en sich die FDR_B nur noch wenig ändert, ein Phänomen, das auch MUCCIARDI (1971) beobachtet hat. Auch FDR_M tritt hier in eine stabile Phase ein, so daß - was die Anzahl der Fehldiagnosen angeht - schon 5/9 der U-en zur (bestmöglichen) Diagnose ausreichen. In dieser stabilen Phase reduziert die mDA die FDR der eDA durchschnittlich um etwa 10 % (von 15 % auf 5 %). Dabei wird etwa die Hälfte der Fehldiagnosen der eDA durch die mDA verbessert!

Vergleicht man die Punkte auf der durchgezeichneten Linie mit den übrigen Ergebnissen, so sieht man, daß jene tatsächlich ausnahmslos niedriger als diese liegen (vgl. Kap. 4.1.).

Außerdem erkennt man, daß die theoretisch (nach Gl. (2-4)) bestimmte Schranke für FDR_M (hier 40 %) sehr grob ist, was durch die Größe von $N = 4$ bedingt ist. (In Kap. 2.4. liegt für den Fall $N = 2$ die berechnete Schranke viel näher an den tatsächlichen Ergebnissen).

1) Auch MICHAELIS (1972) stellte bei Anwendung einer Art von mDA - die Krankheiten mit den beiden größten a posteriori Wahrscheinlichkeiten werden als möglich betrachtet - fest, daß ein kleineres M_a eine geringere FDR erbringt, wertete dies jedoch als "Zufallseffekt".

2.3.2. Verlust an diagnostischer Schärfe

Dieses Kriterium beschreibt die nachteiligen Auswirkungen der mDA. Hier wird gezeigt, für wieviele Fälle die DA un-
nötigerweise modifiziert wird (Q_n) und wie groß der durchschnittliche Verlust an Information gegenüber der eDA (ΔINF) wird (Diagramm 2.2). Beide Maße für den VaDS zeigen einen ähnlichen Verlauf. Sie nehmen monoton mit wachsender Merkmalsanzahl ab, anfangs sehr schnell, später langsamer (bes. ΔINF). Dabei werden zunächst 85 % der DA echt modifiziert, geht ca. 65 % der maximal möglichen Information $TINF_E$ ($= 1.363$) verloren. Schließlich, bei Benutzung aller neun U-en beträgt aber der Anteil der echt mDA nur noch ca. 20 %, der VaDS von ca. 0.1 läßt nur noch 7 % von $TINF_E$ verloren gehen.

Auch hier tritt ΔINF ab $M_a = 6$ in eine einigermaßen stabile Phase ein, während Q_n weiter abnimmt. Außerdem stellt man fest, daß die drei eingezeichneten Punkte für andere AM wieder ungünstiger liegen als die der "besten" Auswahl (vgl. Kap. 4.1.).

Ist die Größe Q_n im Vergleich zu ΔINF auch diejenige, die eine einfachere Interpretation zuläßt, so ist nur ΔINF in der Lage, den VaDS auch bei der Betrachtung des Einzelfalles zu messen (vgl. Kap. 3.2.). Dieses stellt unseres Wissens den ersten Versuch dar, ein Maß für das Differenzierungsvermögen einer DA zu konstruieren. Daß ein solches Kriterium als wünschenswert und tatsächlich notwendig erachtet wird, zeigt noch das KNILL-JONES (1975) entnommene Beispiel in Kap. 2.4..

Haben wir bis jetzt Vor- und Nachteile der mDA getrennt abgehandelt, wollen wir nun ein Kriterium vorstellen, daß beide Gesichtspunkte beinhaltet und so eine Gütebeurteilung der mDA ermöglicht.

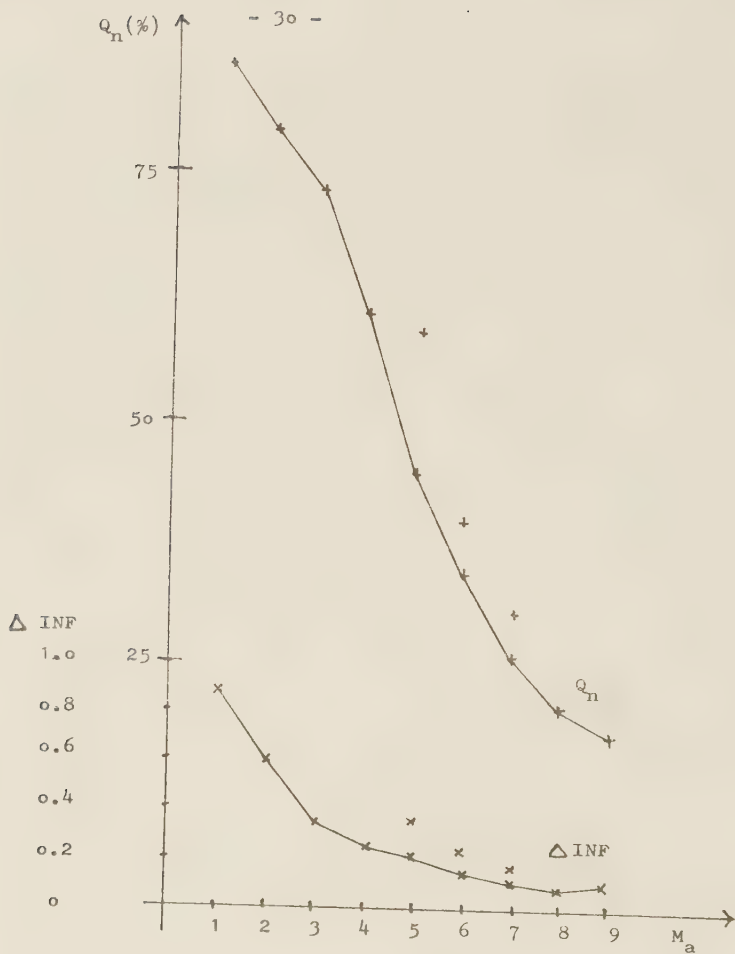


Diagramm 2.2.

Durch die mDA erlittener Verlust an diagnostischer Schärfe, ausgedrückt durch die Kriterien Q_n (+) und ΔINF (x) für die jeweils "beste" U-Auswahl (durchgezeichnete Kurven) und einige andere AM (Punkte) bei Variation der U-Anzahl M_a und Anwendung des NSV.

Beide Maße für den VaDS nehmen mit zunehmendem M_a monoton ab, wobei die Abnahme für kleines M_a besonders groß ist.

2.3.3. Güte der mDA: Der Modifikationsquotient (MQ)

Die Güte der mDA wird, wie erwähnt, durch das Verhalten ihrer Vor- und Nachteile gegenüber der eDA bestimmt. Dabei spielen einerseits

- die relative Reduktion der FDR der eDA

$$Q_{FDR} = (FDR_B - FDR_M) / FDR_B , \text{ andererseits}$$

- der relative Verlust an diagnostischer Schärfe

$$(Q_{INF} = \frac{\Delta_{INF}}{T_{INF}} \quad \text{bzw. die Größe } Q_n)$$

eine Rolle. Wie sich diese Werte bei verschiedener U-Anzahl verhalten, zeigt Tab. 2.1.. Dabei wurde wieder für das jeweilige M_a das " beste " NSV gewählt.

Tab. 2.1.:

M_a	1	2	3	4	5	6	7	8	9
$Q_{FDR} (\%)$	97,4	94,1	95,2	88,2	63,6	73,3	54,5	50,0	36,2
$Q_{INF} (\%)$	64,8	45,0	26,6	23,7	16,4	13,5	9,1	6,1	6,9
$Q_n (\%)$	85	78	72	40	44	34	26	20	18

Relative Reduktion der FDR (Q_{FDR}) und relativer Verlust an diagnostischer Schärfe (Q_{INF} , Q_n) in Abhängigkeit der der U-Anzahl M_a . Q_{FDR} ist bei gleichem M_a größer als Q_{INF} bzw. Q_n . Ein guter Kompromiß zwischen niedriger FDR und großer DS ergibt sich bei $M_a = 5$.

Wie schon wegen theoretischer Überlegungen (Gl. (2-7)) zu erwarten, ist die relative Reduktion der FDR_B stets (meist erheblich) größer als der Verlust an diagnostischer Schärfe. Ein günstiger Kompromiß zwischen großem Q_{FDR} und niedrigem Q_{INF} ergibt sich dabei am Beginn der bereits mehrfach erwähnten stabilen Phase ($M_a = 5$, $M_a = 6$). Diese Merkmalskombinationen haben also nicht nur den Vorteil, unnötige U-en zu vermeiden, sondern stellen auch die U-Auswahl dar, die - verglichen mit dem Verlust an diagnostischer Schärfe - durch Anwendung der mDA die FDR der eDA relativ am meisten verringert.

Bei der praktischen Anwendung des NSV stellt sich aber das Problem, diese Grenze - den Beginn der stabilen Phase - v o r dem Testen aller möglichen NSV-en festzustellen. Dazu sei schon jetzt erwähnt, daß die durchschnittliche U-Anzahl der in Kap. 3 entwickelten sequentiellen Selektionsverfahren (SV) fast immer an dieser Grenze liegen. So wird bei Anwendung dieser neuen Methode der günstige Kompromiß zwischen Q_{FDR} und Q_{INF} i.a. automatisch erreicht, meist sogar noch verbessert!

Eine Möglichkeit, Vor- und Nachteile der mDA in e i n e m Kriterium zueinander in Beziehung zu setzen, ergibt sich aus der Analyse der Rate Q_n der echt modifizierten DA. Q_n setzt sich ja aus der Rate der durch die mDA korrigierten Fehldiagnosen der eDA, $\Delta FDR = FDR_E - FDR_M$ und der Rate der unnötigerweise mDA $Q_W = Q_n - \Delta FDR$ zusammen. Der aus diesen beiden Größen gebildete M o d i f i k a t i o n s q u o t i e n t $MQ = Q_W / \Delta FDR$ gibt dann an, wieviele Ekgs durchschnittlich durch die mDA unnötigerweise einer reservierten Beurteilung zugeführt werden, um eine Fehldiagnose der eDA zu korrigieren.

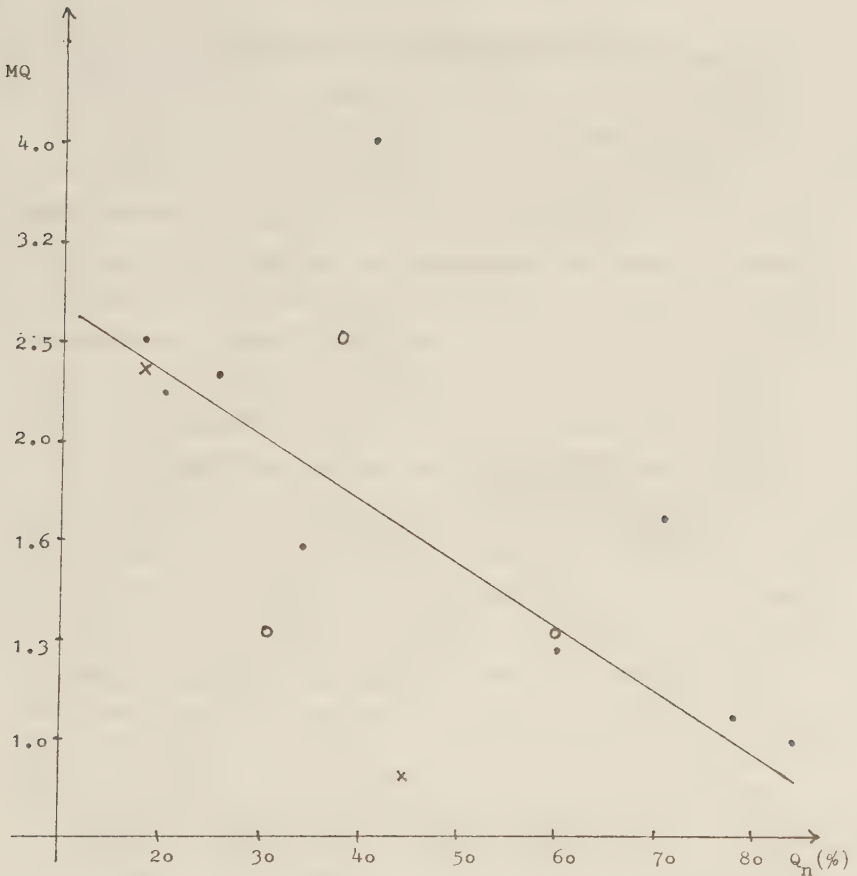


Diagramm 2.3.

Modifikationsquotient in halblogarithmischer Darstellung in Abhängigkeit der Rate echt mDA-en (Q_n) für das jeweils "beste" NSV (•, gemittelt durch den Kurvenzug), andere NSV-en (○) und die Ergebnisse für die DD des Ikterus (x, Kap. 2.4.). Mit wachsendem Q_n sinkt die Rate derjenigen Ekgs, die man unnötigerweise reserviert beurteilen muß, um eine Fehldiagnose der eDA durch die mDA zu korrigieren (MQ). Dabei ergibt sich für das "beste" NSV im halblogarithmischen Maßstab eine lineare Beziehung.

Er ist somit - als eine Art Kosten-Nutzen-Quotient - die entscheidende, und auch sehr anschauliche Größe, um die Konsequenzen einer mDA zu erfassen.

Beobachtet man MQ in halblogarithmischer Darstellung in Abhängigkeit der insgesamt getätigten echt mDA, bekommt man Diagramm 2.3. Dabei ist auffallend, daß $\ln (MQ)$ zu Q_n bei Anwendung der für jedes feste M_a "besten" NSV annäherungsweise in linearer Beziehung steht. Die übrigen Punkte, Ergebnisse anderer NSV und der in 2.4 diskutierten DD des Ikterus, liegen teils unter, teils über der eingezeichneten Gerade.

Bei Anwendung der jeweils "besten" NSV verbessert sich mit wachsender Zahl echt mDA also das Verhältnis von unnötigerweise mDA-en zu korrigierten mDA-en. Erreicht schließlich der Anteil echt mDA an allen DA insgesamt ca. 80 %, muß man, um eine Fehldiagnose der eDA mit der mDA zu verhindern, durchschnittlich nur ein anderes Ekg reserviert beurteilen. Dieser Fall, der der Benutzung lediglich einer U ($M_a = 1$) entspricht, leidet aber unter einem großen VaDS (vgl. Kap. 2.3.2.). In dem uns hauptsächlich interessierenden Intervall zwischen 20 % und 30 % - dies entspricht $5 \leq M_a \leq 7$ - kommen im Durchschnitt auf eine korrigierte etwa 1,5 bis 2 unnötigerweise mDA.

Neben der Abhängigkeit von der Rate Q_n der echt mDA zeigt der MQ aber auch eine Änderung bei Variation des U-Verfahrens. So werden die sequentiellen Verfahren in Kap. 3 - besonders diejenigen mit individueller Merkmalsauswahl - einen besonders niedrigen MQ ergeben, gleichzeitig aber auch eine ähnliche, wenn auch nicht so strikte, Abhängigkeit von Q_n zeigen.

2.4. Ein zweites Beispiel: Die Differentialdiagnose des Ikterus

In ähnlicher Weise wie bei der hier gezeigten DD kongenitaler Herzvitien hat KNILL-JONES (1973, 1975) die DD des Ikterus durchgeführt. Er unterschied dabei zwischen 9 konservativ (" Medical Cases ") und 2 chirurgisch zu behandelnden K-en (" Surgical Cases ", Tab. 2.2.).

Tab. 2.2.

"Medical Cases"	"Surgical cases"
1. Akute Virushepatitis	1. Tumor der extrahepa-
2. Fulminante Hepatitis	tischen Gallenwege
3. Leberzirrhose	2. Chole- und Choledocholithiasis
4. Chronisch aktive Hepatitis	
5. Primär biliäre Zirrhose	
6. Lebermetastasen	
7. Ikterus durch Chlorpromazin	
8. Ikterus durch andere Medikamente oder Drogen	
9. M. Gilbert	
<u>11 diagnostische Kategorien für die DD des Ikterus</u> (nach KNILL-JONES, 1975)	

Die benutzten U-en (insgesamt 102) entnahm er der Anamnese (45), der körperlichen U (12), Laborresultaten (34) und der Leberszintigraphie (11). Diese große Zahl macht auch ersichtlich, warum er für die Berechnung der WK-en $q (j | x^{(AM)})$ nur diskrete univariate WK-Verteilungen verwendete (obwohl gerade bei großer U-Anzahl multivariate Methoden sehr nützlich sind). Die für die Bestimmung der Funktionen $f (x^{(AM)} | j)$ notwendige

Kenntnis der tatsächlichen Diagnosen der Probanden eines "Trainingskollektivs" (vgl. Kap. 1.2.2.) ermittelte KNILL-JONES mit Hilfe von Leberbiopsie- und Laparatomie-Resultaten. Für zwei verschiedene Situationen führte er dann eine mDA durch:

(A) Unterscheidung "medical case" - "Surgical Case"

Für diesen 2-Klassen-Fall ordnete er ein Muster x (die U-Anzahl M_a wird hier außer acht gelassen) dann genau zu ("certain cases"), falls ein $q(j|x) \geq 0.96$ ($j=1,2$) war. Alle anderen Muster waren "uncertain" - in unserer Terminologie: echt mDA (vgl. Kap. 2.2.3.). Mit unserer Bezeichnungsweise bekam KNILL-JONES so folgende Ergebnisse:

80% der Fälle waren "certain", so daß 20% echt mDA waren: $Q_n = 20\%$. Da dieses totale Zurückweisen¹⁾ auch totalen Informationsverlust bedeutet ($\inf(\delta_M(x)) = 0$), war auch $\Delta \text{INF} / \text{TINF}_E = 20\%$. Von den "certain cases" wurden 6% falsch zugeordnet, so daß $\text{FDR}_M = 0.8 \cdot 6\% = 4.8\%$ erreicht wurde, was wegen $c_j = 4\%$ - dies entspricht der Forderung $q(j|x) \geq 0.96$ - und der sich hieraus ergebenden Schranke für FDR_m von $\bar{c} = 8\%$ (vgl. Kap. 1.2.2.) auch ungefähr erwartet werden durfte.²⁾ Bei eindeutiger Zu-

-
- 1) Die theoretischen Eigenschaften eines solchen "Reject" Verfahrens sind von HESS (1979 A) beschrieben worden.
 - 2) Für den Fall $N=2$ erfüllt die mDA nach QUESENBERRY und GESSAMAN ein Optimalitätskriterium: Von allen möglichen Entscheidungsregeln, für die $\text{FDR}(j) \leq c_j$ gilt, ist die mDA nach Gl. (2-1) diejenige, für die Q_n minimal wird.

ordnung (entsprechend Gl. (1-7) bekam KNILL-JONES $FDR_B = 11\%$, so daß die mDA die FDR um ca. 6% verringerte, aber 14% aller Fälle unnötigerweise reserviert beurteilte, was einen MQ von 2.3 ergab (Tab. 2.3., Teil (A)).

(B) DD aller 11 Krankheiten

Hier benutzte er nur die U-en, die innerhalb einer 48-Stunden-Frist nach Aufnahme in die Klinik verfügbar waren. Wieder unter Anwendung von $c_j = 4\%$ bekam er die in Teil (B) von Tab. 2.3. aufgeführten Werte.

Tab.2.3.

	N	FDR_B	FDR_M	Q_n	Q_w	FDR	MQ	$\Delta INF/TINF_E$
(A)	2	11%	5%	20%	14%	6%	2.3	20%
(B)	11	29%	5%	45%	21%	24%	0.9	?

Güteparameter der eDA und der mDA: Auch für die DD des Ikterus folgen sie im wesentlichen den in Kap.2.3. beschriebenen Gesetzmäßigkeiten

Dabei fällt auf, daß die FDR der mDA trotz der erheblich größeren K-Anzahl und der damit verbundenen größeren Schranke für FDR_M ($\bar{c} = 44\%$) wie bei $N=2$ nur 5% beträgt, wogegen sich die FDR der eDA von 11% auf 29% erhöht. Allerdings macht es die Anwendung der mDA für $N=11$ nötig, fast die Hälfte aller Fälle (45%) acht mDA-en zuzuführen, was - entsprechend Diagramm 2.3. - einen niedrigen MQ zur Folge hat.

Somit folgen die hier erläuterten Ergebnisse der DD des Ikterus qualitativ den Gesetzmäßigkeiten, die am Beispiel der DD kongenitaler Herzvitien aufgezeigt wurden. Jedoch

konnte KNILL-JONES, obwohl auch er in dieser 11-Klassen-Situation für die "uncertain cases" eine mDA in Form einer neuen DDA traf, den daraus resultierenden Informationsverlust nicht quantifizieren (" In the uncertain cases a differential diagnosis is given, and this cannot easily be expressed as an overall accuracy figure "), so daß die entsprechende Größe $\Delta \text{INF}/\text{TINF}_E$ nicht bestimmt werden konnte. Dieses Zitat unterstreicht jedoch noch einmal die Notwendigkeit und Nützlichkeit der in Kap. 2.2.3. eingeführten Größe ΔINF .

2.5. Schlußfolgerungen

Zusammenfassend ist festzustellen, daß die mDA in vielen Fällen eine hervorragende Alternative zu der eDA darstellt. Sie ist häufig in der Lage, die FDR der eDA erheblich zu reduzieren, ohne einen zu großen VaDS zu riskieren. Ob bei Anwendung der mDA aber einer niedrigen FDR_M oder einem geringen VaDS der Vorzug zu geben ist, muß bei jeder DD-Fragestellung neu geklärt werden. So wird man, wenn man für ein DD-Problem noch über den vorhandenen U-Pool hinaus wesentliche, nicht Computer-compatible U-en zur Verfügung hat, eher einen großen VaDS eingehen wollen als eine große FDR - was durch die Vorgabe kleiner Klassifikationskonstanten c_j zu realisieren ist. In jedem Fall ist also die mDA immer im Zusammenhang des gesamten Diagnoseverlaufs zu sehen und zu variieren. Stets stellt die Gütebeurteilung mit Hilfe des MQ dabei eine geeignete Methode dar, Kosten und Nutzen der mDA einzuschätzen.

Besonders bei der Anwendung geringer U-Zahlen - und dies stimmt mit Erfahrung ärztlichen Diagnoseverhaltens überein - ist die mDA eine ausgezeichnete Lösung. Dagegen ist ihr Gewinn bei sehr hohen U-Zahlen zweifelhaft. Sehr günstig schneidet die mDA in unserem Beispiel der DD kongenitaler Herzvitien bei einer U-Anzahl von etwa $2/3$ der maximal möglichen U-en ab. In dieser Situation - zu Beginn einer in jeder Hinsicht stabilen Phase der DF - scheint sie einen "optimalen" Kompromiß zwischen Reduktion der FDR und Erhaltung des diagnostischen

Differenzierungsvermögens herzustellen. Dieser Übergang in die stabile Phase der DF ist jedoch vorab bei Anwendung der NSV-en kaum zu bestimmen. Dagegen wird er bei den in Kap. 3 eingeführten sequentiellen Verfahren (SV) automatisch erreicht, so daß dann der Nutzen der mDA noch deutlicher wird.

Die Erkenntnis, daß für eine gute DA schon ein geringer Teil der möglichen U-en ausreicht, hat auch KNILL-JONES (1975) bei der DD des Ikterus formuliert (bei ihm genügte ein Drittel der maximalen U-Anzahl). An dieser Stelle sei schon erwähnt, daß er seine U-en schrittweise durchführte und individuell auswählte. Die Anwendung solcher SV-en im nächsten Kapitel wird zeigen, daß sie in der Lage sind, die diagnostische Potenz aller anderen Verfahren noch zu übertreffen.

KAPITEL III

SEQUENTIELL INDIVIDUELLE UNTERSUCHUNGS-AUSWAHL

3.1. Einleitung

Ein sequentielles, schrittweises Vorgehen bei der DF ist i.a. für das ärztliche Diagnoseverhalten kennzeichnend (vgl. Kap. 1.1.). Auch in der CDF wurden daran anlehnend Methoden entwickelt, deren Anwendung jedoch so große FDR-en ergab, daß sie bald von anderen, nicht-sequentuellen Verfahren (NSV) abgelöst wurden. So gelten etwa die Ausführungen von MUCCIARDI (1971) über die optimale Auswahl von Ekg-Parametern nur für solche NSV-en. Diese, meist von Resultaten der mathematischen Entscheidungstheorie ausgehenden "multivariaten" Entscheidungsregeln sind in Kap. 1.2. vorgestellt und in Kap. 2 modifiziert worden. Allerdings ist schon von einigen Autoren festgestellt worden, daß bei der Durchführung von NSV-en viele U-en redundant (BAY, 1976) oder bei der Betrachtung des individuellen Falls sogar störend wirken können (KNILL-JONES, 1975). Diese beiden Tatsachen sind auch in Kap. 1.3. anhand von Beispielen aufgezeigt worden; außerdem wurde hier angedeutet, daß ein gutes s e q u e n t i e l l e s Diagnoseverfahren (SV) bestimmt wird durch eine geeignete U-Auswahl und durch eine rechtzeitige Beendigung der DF mit Hilfe einer Stoppregel (SR). Beide müssen so beschaffen sein, daß sie den unterschiedlichen Bedingungen in jedem konkreten Fall Rechnung tragen. So sollen sie insbesondere (vgl. (I) - (IV) aus Kap. 1.1.) folgende Forderungen erfüllen:

- Sprunghafte Diagnoseverläufe mit häufig wechselnder DDA sollen durch eine geeignete Reihenfolge der U-en vermieden werden.
- Bei schon früh richtigen und eindeutigen DA-en sollen weitere U-en unterlassen werden, um Zeit und Kosten zu sparen und eine Belästigung oder Gefährdung des Patienten weitgehendst zu vermeiden.
- Nicht eindeutige DA-en sollen so lange weiter abgeklärt werden, bis sie ausreichend differenziert sind.
- Einer Entwicklung der DF in die falsche Richtung soll durch rechtzeitigen Stopp vorgebeugt werden.

Diese Forderungen können in optimaler Weise nur erfüllt werden, wenn U-Auswahl und SR

- die bei jedem Schritt der DF erhobenen neuen DD-Erwägungen (DDA) in die Auswahl der nächsten U bzw. die Beendigung des Verfahrens mit einbeziehen (i n d i v i d u e l l e Verfahren) und
- die multidimensionalen Abhängigkeiten der U-en untereinander durch multivariate Analyse berücksichtigen (m u l t i v a r i a t e Verfahren).

Als Maßstäbe für die Güte der verwendeten SV-en bieten sich gemäß der Erörterungen in Kap. 2

- bei Anwendung der eDA FDR_B und
- bei Anwendung der mDA FDR_M einerseits und der VaDS, repräsentiert durch ΔINF und Q_n , andererseits an.

Weitere Indikatoren sind der Verlauf des MQ und die Anzahl der durchschnittlich benötigten U-en.

3.2. Methode

3.2.1. Theoretische Grundlagen

Das wichtigste Problem bei der Auswahl geeigneter U-en ist, bei jedem Schritt der DF im v o r a u s abschätzen zu können, welche U zur Abklärung einer verbliebenen DDA am besten geeignet ist. Dazu muß man wissen, welches Risiko einer Fehldiagnose man bei Anwendung einer geplanten U - unter Berücksichtigung bereits erzielter Ergebnisse - zu erwarten hat. Diese zu erwartende FDR für alle Konstellationen der DDA und für alle möglichen Kombinationen von U-en anhand von Testkollektiven zu prüfen, ist schlechterdings unmöglich. HESS (1979A, B) gelang es jedoch, die FDR geplanter U-en schon nach Kenntnis der relativen Häufigkeiten p_j der DD in Frage kommenden K-en und ihrer multivariaten Parameter μ_j und Σ_j (vgl. Kap. 1.2.2.) in guter Näherung zu bestimmen. Dazu ist es nur nötig, die genannten Größen für e i n e DDA (a l l e N K-en) und für e i n e Kombination von U-en (a l l e M Ekg-Merkmale) anhand eines Trainingskollektivs zu berechnen - dies sind genau dieselben Voraussetzungen, die zur Durchführung der "multivariaten Diskriminanzanalyse" als NSV erfüllt sein müssen.

Als Hilfsgrößen werden dabei die "Bhattacharyya-Koeffizienten" (BK) benutzt, die sich bei Anwendung der aktuellen U-en AM wie folgt berechnen lassen:

$$\begin{aligned} \phi_{ij}^{(AM)} = & \left| \det(\Sigma_i^{(AM)}) \cdot \det(\Sigma_j^{(AM)}) \right|^{1/4} \cdot \det(\Sigma_{ij}^{(AM)})^{-1/2} \\ & \cdot \exp \left\{ -\frac{1}{4} \left(\mu_i^{(AM)} - \mu_j^{(AM)} \right)^T \left(2 \Sigma_{ij}^{(AM)} \right)^{-1} \left(\mu_i^{(AM)} - \mu_j^{(AM)} \right) \right\} \\ & \text{Gl. (3-1)} \end{aligned}$$

mit $\Sigma_{ij}^{(AM)} = (\Sigma_i^{(AM)} + \Sigma_j^{(AM)})/2$ für $i \neq j = 1, \dots, N$

Die BK β_{ij} erlauben es nun, die bei Benutzung von AM zu erwartende FDR_B nur bei Kenntnis von p_j , $\mu_j^{(AM)}$ und $\Sigma_j^{(AM)}$ abzuschätzen:

$$\begin{aligned} & \sum_{i=1}^N \sum_{j=1, < i}^N p_i p_j \beta_{ij}^2(AM) \\ & \leq \frac{N-1}{N} \left(1 - \sqrt{1 - \frac{2N}{N-1} \cdot \sum_{i=1}^N \sum_{j=1, < i}^N p_i p_j \beta_{ij}^2(AM)} \right) \\ & \leq FDR_B(AM) \leq \sum_{i=1}^N \sum_{j=1, < i}^N \sqrt{p_i p_j} \beta_{ij}(AM) \quad \text{Gl. (3-2)} \end{aligned}$$

Wie der folgende Abschnitt zeigen wird, stellen diese Abschätzungen den Kern eines jeden SV dar.

3.2.2. Multivariat Individuell Sequentielles Selektionsverfahren

Wie sich die in der Einleitung zu diesem Kapitel formulierten Forderungen an ein gutes SV mit Hilfe der theoretischen Überlegungen aus Kap. 3.2.1. realisieren lassen, zeigt Diagramm 3.1.. Hier ist das Procedere des "Multivariat Individuell Sequentiellen Selektionsverfahrens" in Form eines Flußdiagrammes aufgezeichnet. In insgesamt 15 Teilen besteht es aus:

- (V) Die relativen Häufigkeiten p_j der DD in Frage kommenden K-en werden vorgegeben. Zusätzlich seien für M Ekg-U-en (maximaler U-Pool PM) die Mittelpunktsektoren μ_j und die Kovarianzmatrizen Σ_j für jede K

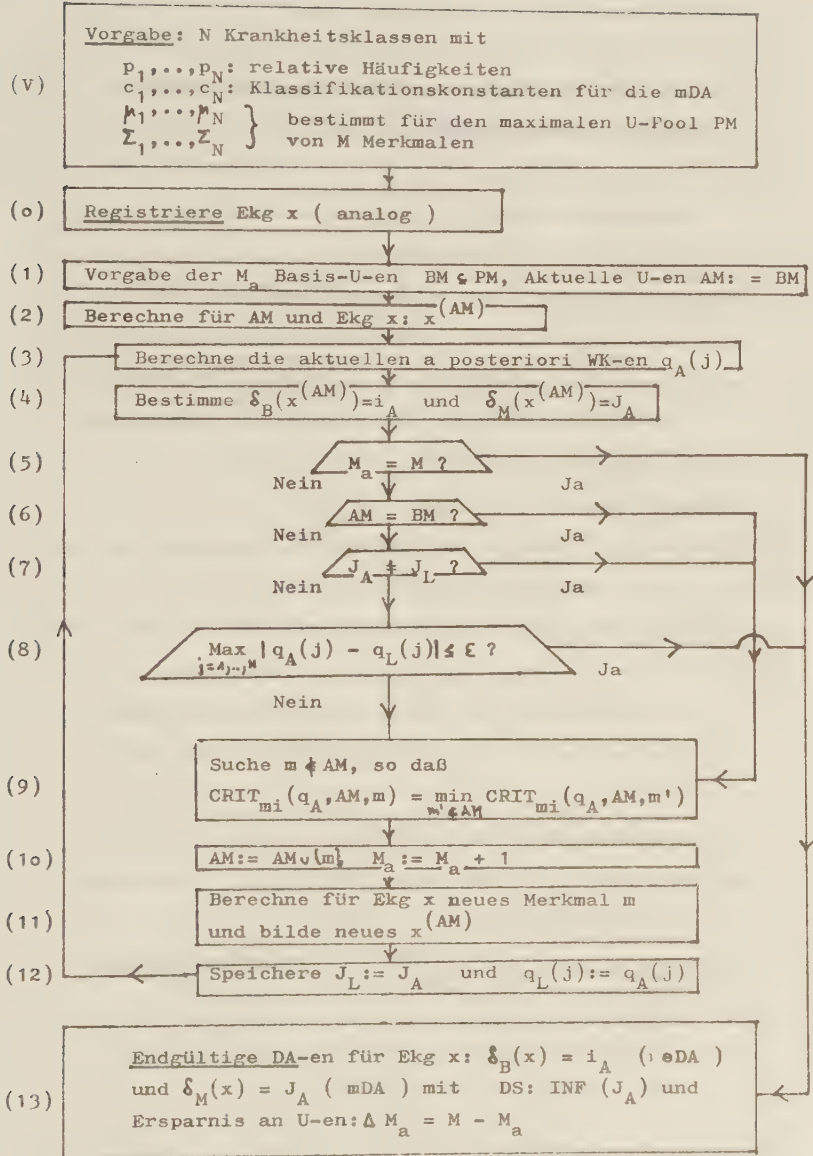


Diagramm 3.1.: Flußdiagramm für die Durchführung des misV

j anhand eines Trainingskollektivs bestimmt worden.
Die für die Durchführung der mDA notwendige Vorgabe der Klassifikationskonstanten c_j (nach Gl. (2-1) soll gewährleistet sein.

- (o) Nach der analogen Registrierung des zur DD anstehenden Ekg's x wird
- (1) anhand von M_a vorgegebenen Basismerkmalen BM aus PM - diese bilden vorläufig die aktuellen Merkmale AM -
 - (2) der M_a -dimensionale aktuelle Ekg-Vektor $x^{(AM)}$ gebildet.¹⁾
 - (3) Mit Hilfe von $x^{(AM)}$ werden die aktuellen a posteriori WK-en $q_A(j) = q(j | x^{(AM)})$ jeder K_j nach Gl. (1-1) berechnet. Dazu ist es nötig, μ_j und Σ_j auf die aktuelle Größe gemäß AM zu reduzieren (nach Kap.1.2.2.).
 - (4) Damit ist es möglich, die aktuelle DA $\delta_B(x^{(AM)}) = i_A$ (eDA, Gl.(1-7)) und $\delta_M(x^{(AM)}) = J_A$ (mDA, Gl.(2-1)) zu tätigen.

- (9) Den wichtigsten Teil des Verfahrens stellt die Auswahl der nächstbesten U_m dar. Es soll dasjenige Merkmal m als nächstes ausgewählt werden, das, zusammen mit AM, die niedrigste - durch $q_A(j)$ gewichtete - FDR erwarten läßt. In Anlehnung an Gl.(3-2) wird verlangt, daß das neue Merkmal m

$$CRIT_{mi}(q_A, AM, m') = \sum_{i=1}^N \sum_{j=1, < i}^N \sqrt{q_A(i)q_A(j)} \beta_{ij}(AM, m')$$

$$\text{minimiert: } CRIT_{mi}(q_A, AM, m) = \min_{m' \notin AM} CRIT_{mi}(q_A, AM, m')^2$$

-
- 1) Die hierfür notwendige Berechnung digitaler Ekg-Merkmale stellt i.a. den aufwendigsten Teil des Verfahrens dar.
 - 2) Der Index "mi" bedeutet: multivariat individuell

Dabei gewährleistet die Abhängigkeit von CRIT_{mi} von q_A die Berücksichtigung des individuellen Falls, die von AM die multidimensionale Abhängigkeit der U-en untereinander.¹⁾

- (10) Anschließend wird die Menge AM um das neue Merkmal m vergrößert, M_a um 1 heraufgesetzt.
 - (11) Nach Berechnung des neuen Merkmals für das Ekg x und
 - (12) Speicherung der letzten mDA ($J_L = J_A$) und der letzten WK-en $q_A(j)$ ($q_L(j) = q_A(j)$)
 - (3) werden für den neu gebildeten Ekg-Vektor $x^{(AM)}$ die aktuellen $q_A(j)$ und DA-en $\mathcal{S}_B(x^{(AM)})=i_A$ und $\mathcal{S}_M(x^{(AM)})=J_A$ bestimmt.
- Diese Vorgehensweise wird solange fortgeführt, wie es die SR gestattet. Das Verfahren wird beendet, falls
- (7) die aktuelle mDA J_A mit der letzten mDA J_L übereinstimmt und gleichzeitig
 - (8) die aktuellen und letzten WK-en $q_A(j)$ und $q_L(j)$ sich höchstens um eine vorgegebene Größe ε (hier $\varepsilon = 0.05$) unterscheiden.
 - (5) Außerdem wird die DF beendet, wenn der PM erschöpft ist, aber
 - (6) sie wird solange fortgesetzt, daß mindestens eine U über BM hinaus ausgeführt wird.
 - (13) Die endgültige DA kann dann in Form der eDA ($\mathcal{S}_B(x)=i$) oder als mDA ($\mathcal{S}_M=J$) erfolgen, die DS wird durch $\text{INF}(\mathcal{S}_M(x))$ (Gl. (2-5)), die gegenüber der maximal

1) Nach (2) ist (9) der aufwendigste Teil des Verfahrens. (9) benötigt - wegen der Berechnung von $(M-M_a)$ BK-en und dem damit verbundenen Aufwand an Matrizenmanipulationen (Gl.(3-1)) - erhebliche Rechenzeit.

möglichen Zahl von M U-en ersparte U-Anzahl durch Δ^M_a gegeben.

Zur Vereinfachung der recht aufwendigen Berechnung von CRIT_{mi} gibt es drei Möglichkeiten der Einschränkung:

md: Die Reihenfolge der U-en wird - unabhängig vom individuellen Fall - vorab bestimmt. Die durch - s c h n i t t l i c h beste Abfolge erhält man durch Minimierung des Kriteriums

$$\text{CRIT}_{md}(AM, m') = \sum_{i=1}^N \sum_{j=1, < i}^N \sqrt{p_i p_j} \beta_{ij}(AM, m').$$

Da AM in die jeweilige Auswahl einbezogen wird, handelt es sich um eine m u l t i v a r i a t e Selektion.

ui: Hier werden die aktuellen U-en AM nicht berücksichtigt, wohl aber die letzte DDA (durch $q_A(j)$: i n d i v i d u e l l):

$$\text{CRIT}_{ui}(q_A, m') = \sum_{i=1}^N \sum_{j=1, < i}^N \sqrt{q_A(i) q_A(j)} \beta_{ij}(m')$$

Dies hat den Vorteil, nur u n i v a r i a t e BK-en berechnen zu müssen, was keine Matrizenmanipulationen erfordert.

ud: Dies stellt eine Kombination von md und ui dar:

$$\text{CRIT}_{ud}(m') = \sum_{i=1}^N \sum_{j=1, < i}^N \sqrt{p_i p_j} \beta_{ij}(m')$$

Da die Abfolge der U-en vorab, und damit nur einmal bestimmt wird, ist der Gewinn an Rechenzeit nur gering.

Auch die SR kann anders gestaltet werden, z.B., wenn man auf Punkt (8) verzichtet und das Verfahren schon stoppt,

wenn zwei aufeinanderfolgende mDA-en übereinstimmen. Die extreme Variation der SR stellen die NSV-en dar, die sofort nach Berechnung der BM die endgültige DA treffen. Der Kern des im Flußdiagramm 3.1 vorgestellten miSV besteht aus zwei einfachen Überlegungen:

- Sind auf einer aktuellen Stufe der DF $q_L(j)$ die zuletzt bestimmten WK-en der K j, so sollen die neuen $q_A(j)$ so beschaffen sein, daß sie eine weitere Differenzierung der DDA so weit wie möglich verwirklichen:
 $|q_A(j) - q_L(j)|$ soll für alle $j=1, \dots, N$ maximal sein.
- Gestoppt werden soll die DF dann, wenn sogar die U, die diese Differenzen maximiert, sie nicht mehr wesentlich verändern kann: $|q_A(j) - q_L(j)| < \epsilon$.

Gestoppt wird also dann, wenn die DF, durch die U-Auswahl zu größter Progredienz getrieben, eine stabile Phase erreicht hat. Diese Berücksichtigung des Diagnoseverlaufs durch die SR stellt neben der mi-U-Auswahl das zweite völlig neue Element in der CDF dar. Die wenigen in der Literatur zu findenden SV-en vernachlässigen nämlich alle die hier geforderte multivariate U-Auswahl. Eine individuelle Selektion nehmen zwar auch ALPEROVITCH (1977) und KNILL-JONES (1973, 1975) in ihren Methoden vor (beide für die DD des Ikterus), dieser nach einem nicht genauer definierten " expected gain of information ", so daß eine Beziehung des SV zur Minimierung der FDR auch hier nicht zu erkennen ist. Die von KNILL-JONES angewendete SR berücksichtigt nur eine Beendigung der DF für " certain cases " (vgl. Kap. 2.4.), nicht aber die Stabilität

des Diagnoseverlaufs. Für dieses Verfahren verglichen STERN et al. (1974) auch die Reihenfolgen der U-en nach der Computerauswahl mit denen von Ärzten. Dabei stellten sie fest, daß diese außerordentlich unterschiedlich waren - ebenso wie die erzielten FDR-en. Gleichzeitig erkannten sie, daß eine große U-Anzahl M_a die Diagnosegenauigkeit des Computers viel stärker verbesserte als die der Ärzte. Dies liegt wohl daran, daß ein Computer eher in der Lage ist, eine große Menge von U-Ergebnissen zu verarbeiten, was auch ein Grund dafür ist, daß in der Ekg-DD erst die Anwendung der " multivariaten Diskriminanzanalyse " die erwähnten niedrigen FDR-en erbrachte.

Allerdings erkannte man auch hier in einigen Fällen (MICHAELIS, 1972), daß gute Resultate auch bei Reduzierung der U-Anzahl zu erzielen waren. Diese " test - reduction ", die GREMY 1979 auf dem europäischen Kongress für medizinische Informatik als eines der wichtigsten Ziele der CDF bezeichnete und die auch bei PIPBERGER (1973) stets eine große Rolle spielt, wurde aber bei der Ekg-DD immer nur nach durchschnittlichen Kriterien angestrebt. Das hier vorgestellt miSV wird jedoch zeigen, daß es in der Lage ist, eine flexible und sparsame U-Auswahl ohne Vergrößerung, ja sogar mit einer weiteren Reduzierung der FDR vorzunehmen.

3.3. Ergebnisse

3.3.1 Demonstration der SV-en an Einzelfällen

(1) Verschiedene SV-en

Wie schnell der Verlauf der DF bei Anwendung des miSV auf die DD kongenitaler Herzvitien sein kann, zeigt das Beispiel des Ekgs Nr. 60. Wie in Kap. 1.3. beobachten wir dabei die Änderung der a posteriori WK-en $q_A(j)$ und der aktuellen mDA (repräsentiert durch den Entscheidungsvektor E) bei schrittweiser Hinzunahme der nächstbesten U (Tab. 3.1.)

Tab.3.1.

SV	M_a	neue U:Nr	$q_A(j)$ für $j =$				$E(j)$ für $j =$			
			1	2	3	4	1	2	3	4
mi	2	5,6	0 3102	0 1963	0 2145	0 2790	1	1	1	1
	3	10	0 0000	0 0110	0 0000	0 9890	0	0	0	1
	4	9	0 0000	0 0216	0 0000	0 9784	0	0	0	1
ud	2	5,6	0 3102	0 1963	0 2145	0 2790	1	1	1	1
	3	1	0 2722	0 2550	0 0045	0 5010	1	1	0	1
	4	2	0 2745	0 0016	0 0004	0 7246	1	0	0	1
	5	3	0 1920	0 0002	0 0000	0 9978	1	0	0	1
	6	4	0 0132	0 0000	0 0000	0 9862	0	0	0	1
	7	9	0 0132	0 0000	0 0000	0 9850	0	0	0	1

miSV und udSV für Ekg-Nr. 60: Änderung der a posteriori WK-en $q_A(j)$ und der aktuellen mDA, repräsentiert durch den Entscheidungsvektor E ($E(j)=1$ bedeutet: $j \in J_A$) bei schrittweiser Hinzunahme der nach dem jeweiligen Kriterium ($CRIT_{mi}$ bzw. $CRIT_{ud}$) nächstbesten U. (BM: Nr.5,6) Das miSV beendet die DF schon nach 4 von 9 möglichen U-en mit der richtigen Diagnose (VSD), das ud erst nach 7 U-en.

Schon die erste zusätzliche U bringt beim miSV die richtige Diagnose mit WK 0.9. Nachdem die nächste U dies stabilisiert ($|q_L(j) - q_A(j)| < 0.05$), wird die DF gestoppt und

dem Ekg die zutreffende K zugeordnet. Dabei sind im Vergleich zur maximal möglichen Anzahl von U-en (9) 5/9 eingespart worden. Dagegen zeigt das udSV zwar auch einen gleichmäßigen, jedoch wesentlich langsameren Verlauf und verzichtet nur auf die Berechnung von 2 Merkmalen.

Eine andere wichtige Eigenschaft des miSV erkennt man am Beispiel des Ekg Nr. 22 (Tab. 3.2). Hier ist der Verlauf des miSV sogar langsamer als der des udSV. Dies liegt allerdings daran, daß dieses SV schon an einer Stelle, an der noch eine weitere Differenzierung der DDA möglich wäre, die DF abbricht, was auf Kosten der DS geht: die DA wird unnötigerweise echt modifiziert (vgl. Kap. 2.2.3.) Dieser Effekt träte auch bei Ekg-Nr. 60 ein, würde man ε von 0.05 auf 0.06 erhöhen. (dann Stopp des udSV nach $M_a=5$).

Man sieht somit, daß nur das Zusammenwirken der mi-U-Auswahl und der SR eine große Ersparnis an U-en und eine große Trennschärfe erbringt. Daß die geschilderten Beispiele typisch sind, wird noch in Kap. 3.3.2. an Er-

Tab. 3.2.

SV	M_a	neue U:Nr	$q_A(j)$ für $j =$				$E(j)$ für $j =$			
			1	2	3	4	1	2	3	4
mi	2	5,6	0.3987	0.1919	0.3878	0.1096	1	1	1	1
	3	2	0.3202	0.6164	0.0004	0.0630	1	1	0	0
	4	10	0.2254	0.7457	0.0001	0.0287	1	1	0	0
	5	1	0.0870	0.8817	0.0003	0.0309	0	1	0	0
	6	3	0.0957	0.8813	0.0000	0.0230	0	1	0	0
ud	2	5,6	0.3987	0.1919	0.3878	0.1096	1	1	1	1
	3	1	0.3202	0.6164	0.0004	0.0630	1	1	1	1
	4	2	0.2254	0.7457	0.0001	0.0287	1	1	1	1
	5	3	0.0870	0.8817	0.0003	0.0309	1	1	1	1

miSV und udSV für Ekg-Nr. 22 (Bezeichnungen wie Tab. 3.1.): Im Gegensatz zum miSV führt das udSV zu einer (unnötigerweise) echt mDA.

gebnissen mit größeren Fallzahlen gezeigt.

Daneben bewirkt das miSV eine Verstärkung der DF. Wie am Beispiel von Ekg-Nr. 4 deutlich wird, zeigt das miSV einen zwar langsamen, aber gleichmäßigen Verlauf, während der des mdSV wechselhaft und damit für eine Fehldiagnose anfälliger ist. Noch häufiger als beim Vergleich beider multivariater SV-en sieht man diesen Unterschied bei dem von miSV mit uiSV oder udSV.

Tab. 3.3.

SV	M _a	neue U:Nr	q _A (j) für j =				E(j) für j =			
			1	2	3	4	1	2	3	4
mi	2	5,6	0.3420	0.0188	0.3549	0.2843	1	0	1	1
	3	10	0.4099	0.0138	0.4614	0.1149	1	0	1	1
	4	1	0.6947	0.0154	0.1583	0.1316	1	0	1	1
	5	3	0.8406	0.0243	0.0000	0.1352	1	0	0	1
	6	11	0.9105	0.0199	0.0000	0.0696	1	0	0	0
	7	4	0.9214	0.0066	0.0000	0.0720	1	0	0	0
	7	4	0.9214	0.0066	0.0000	0.0720	1	0	0	0
md	2	5,6	0.3420	0.0188	0.3549	0.2843	1	0	1	1
	3	10	0.4099	0.0138	0.4614	0.1149	1	0	1	1
	4	2	0.3730	0.0169	0.5238	0.0863	1	0	1	0
	5	1	0.5890	0.0157	0.3111	0.0842	1	0	1	0
	6	3	0.8607	0.0385	0.0000	0.1008	1	0	0	1
	7	4	0.8940	0.0255	0.0000	0.0805	1	0	0	0
	8	9	0.8705	0.0450	0.0000	0.0845	1	0	0	0
	8	9	0.8705	0.0450	0.0000	0.0845	1	0	0	0

miSV und mdSV für Ekg-Nr. 4 (Bezeichnungen wie Tab. 3.1.):

Während das miSV einen gleichmäßigen Verlauf zeigt, ist der des mdSV wechselhaft und damit für das Entstehen einer Fehldiagnose anfälliger.

(2) Reihenfolge der Untersuchungen

Gemäß der Auswahlkriterien für die verschiedenen SV-en ist die U-Reihenfolge nicht nur von der Wahl des SV abhängig, sondern, wie auch die Ekg-Beispiele in Tab. 3.1 - 3.3 zeigen, für die individuellen SV-en (miSV und

uiSV) auch von dem jeweils untersuchten Ekg-Muster. Auffallend ist jedoch, daß für die Patienten, die tatsächlich dieselbe Krankheit haben, die U-Reihenfolge - bei gleichem SV - in großem Maße konstant ist (Tab. 3.4).

Tab. 3.4.

SV	mi				md	ud	ui			
	Krankheitsklasse						Krankheitsklasse			
	1	2	3	4			1	2	3	4
	10	2	2	10	10	1	1	2	2	1
	1	3	10	1	2	2	2	3	1	2
	3	10	1	4	1	3	3	9	3	4
					3	4				
					4	9				
					9	10				
					11	11				

U-Reihenfolge für die verschiedenen SV-en: Liegt für ein untersuchtes Ekg tatsächlich dieselbe Krankheit zugrunde, so werden für das jeweils gleiche SV die U-en weitgehend in derselben (angegebenen) Reihenfolge ausgeführt. Die Reihenfolge der U bei den durchschnittlichen SV-en - nach Konstruktion unabhängig vom individuellen Fall - ergibt sich als Kompromiß der Reihenfolgen des entsprechenden individuellen SV. Alle Ergebnisse gelten wiederum für die Vorgabe der Basismerkmale Nr. 5 und Nr.6.

Dabei sei wiederum angenommen, daß die schrittweise U-Auswahl ausgehend von den Basis-U Nr.5 und Nr.6 vorgenommen wird. Nimmt man zum miSV die Reihenfolge der U-en beim mdSV zum Vergleich, so sieht man, daß diese sich - nach Konstruktion unabhängig vom jeweiligen Ekg - als Kompromiß der U-Reihenfolgen der einzelnen K-en beim miSV ergibt. Dasselbe gilt für den Vergleich von uiSV und udSV.

Ein weiteres aufschlußreiches Ergebnis ist, daß sich multi- und univariate SV-en in der Abfolge der U-en z.T. fundamental unterscheiden. So fällt auf, daß beim mdSV das Merkmal Nr. 10 als erstes, beim udSV aber erst als vorletztes ausgewählt wird. Zu diesem Phänomen hat schon MICHAELIS (1972) festgestellt, daß es auf die unterschiedliche statistische Abhängigkeit der einzelnen U-en zurückzuführen ist.

ALPEROVITCH (1976) diskutierte für sein univariates SV zur DD eines Rundherdes in der Lunge ebenfalls die Reihenfolge der benutzten U-en und stellte fest, daß sich die U-Auswahl seiner CDF von der der äDF stark unterschied. Insbesondere war die durchschnittliche Anzahl der durchgeführten U-en bei der CDF viel geringer als bei der äDF - Tatsachen , die in Kap. 4.4. bei der Anwendung unserer Ergebnisse auf die äDF noch einmal aufgegriffen werden.

(3) Das miSV für verschiedene Verläufe der DF

In Kap.1.3. ist an fünf Beispielen erläutert worden, welche Forderungen im einzelnen an geeignete Selektionsverfahren und Stoppregeln innerhalb eines SV zu stellen sind. Dort war die U-Auswahl beginnend mit den Ekg-U-en Nr. 5 und 6 nach dem miSV durchgeführt, eine SR aber, um den gesamten Verlauf der DF demonstrieren zu können, nicht angewendet worden. Mißt man das miSV und die Anwendung der mDA an den in Kap. 3.1 formulierten Maßstäben für die genannten Beispiele aus Kap. 1.3., erhält man folgende Resultate (Tab. 3.5.):

Bei Ekg-Nr. 31 wird das miSV nach 4 Schritten und einer U-Anzahl von $M_a=5$ gestoppt. Die bei Anwendung aller 9 U-en sich ergebende Fehldiagnose (vgl. Tab. 1.4.)

wird so verhindert, gleichzeitig wird auf die Ausführung von 4/9 aller möglichen U-en verzichtet.

In einigen Fällen benötigt die Korrektur einer falschen eDA (bei Benutzung aller U-en) nicht nur das miSV, sondern zusätzlich die Anwendung einer mDA. So verhindern beide gemeinsam bei Ekg-Nr. 1 eine Fehldiagnose (vgl. Tab. 1.5.). Die eDA hätte sowohl bei Benutzung von nur 3 U-en (nach dem miSV) als auch bei Benutzung aller 9 U-en diesem Ekg die falsche Krankheit zugeordnet.

Andererseits zeigt Ekg-Nr. 2, daß die Anwendung einer mDA auch unnötigerweise erfolgen kann. Hier würde die eDA für das miSV die richtige Krankheit erkennen, die mDA fügt dieser nur eine weitere (falsche) Krankheit als zu berücksichtigende Diagnose hinzu (vgl. Tab. 1.6.).

In allen drei Beispielen wird außerdem nur ein Teil der maximal möglichen U-en für die DA benötigt. Eine besonders große Ersparnis an U-en erbringen jedoch die problemlosen, schnellen und gleichmäßigen Verläufe der DF, wie von Ekg-Nr. 20 repräsentiert (vgl. Tab. 1.3.). Nur vier U-en reichen zum Erkennen der richtigen Krankheit aus, alle übrigen U-en sind zwar für die Richtigkeit der DA nicht schädlich, wohl aber überflüssig.

In anderen Fällen jedoch ist eine große Anzahl von U-en nötig, um die DDA bis auf eine Krankheitsklasse zu reduzieren. Bei Ekg-Nr. 36 erfordert die schrittweise Differenzierung 8 U-en, 9 U-en sind notwendig, um die Stabilität der DF festzustellen und die SR zur Anwendung zu bringen. (vgl Tab. 1.2.).

Tab. 3.5.

Ekg Nr.	M _a	neue U: Nr	q _A (j) für j =				E(j) für j =			
			1	2	3	4	1	2	3	4
31	2	5,6	0.1167	0.8027	0.0677	0.0129	1	1	0	0
	3	2	0.0722	0.9233	0.0001	0.0043	0	1	0	0
	4	3	0.0100	0.9883	0.0000	0.0017	0	1	0	0
	5	11	0.0083	0.9909	0.0000	0.0007	0	1	0	0
1	2	5,6	0.3327	0.0095	0.0704	0.5874	1	0	0	1
	3	10	0.3609	0.0135	0.0610	0.5645	1	0	0	1
2	2	5,6	0.2412	0.0278	0.0398	0.6911	1	0	0	1
	3	10	0.6411	0.0221	0.1065	0.2303	1	0	1	1
	4	1	0.8038	0.0194	0.0058	0.1710	1	0	0	1
	5	3	0.8039	0.0275	0.0000	0.1636	1	0	0	1
20	2	5,6	0.0172	0.9014	0.0696	0.0119	0	1	0	0
	3	2	0.0005	0.9995	0.0000	0.0000	0	1	0	0
	4	3	0.0001	0.9999	0.0000	0.0000	0	1	0	0
36	2	5,6	0.2499	0.3680	0.2485	0.1335	1	1	1	1
	3	2	0.3021	0.5174	0.0418	0.1386	1	1	0	1
	4	10	0.4362	0.5186	0.0207	0.0245	1	1	0	0
	5	1	0.3663	0.5848	0.0300	0.0189	1	1	0	0
	6	3	0.1808	0.8065	0.0000	0.0128	1	1	0	0
	7	9	0.1245	0.8636	0.0000	0.0119	1	1	0	0
	8	11	0.0431	0.9505	0.0000	0.0064	0	1	0	0
	9	4	0.0197	0.9739	0.0000	0.0063	0	1	0	0
47	2	5,6	0.2499	0.3680	0.2485	0.1335	1	1	1	1
	3	2	0.3589	0.1025	0.3219	0.2167	1	1	1	1
	4	10	0.5126	0.0757	0.3955	0.0162	1	0	1	0
	5	1	0.5325	0.3455	0.1202	0.0018	1	1	1	0
	6	3	0.5488	0.4502	0.0001	0.0008	1	1	0	0
	7	9	0.2153	0.7831	0.0000	0.0015	1	1	0	0
	8	11	0.1092	0.4414	0.4491	0.0003	1	1	1	0
	9	4	0.2681	0.7310	0.0009	0.0000	1	1	0	0

miSV für sechs verschiedene Ekgs (Bezeichnungen wie Tab.3.1.):

- Ekg-Nr.31: Die bei Benutzung aller 9 U-en sich ergebende Fehldiagnose (Tab.1.4.) wird durch den Stopp verhindert.
- Ekg-Nr.1: miSV und mDA vermeiden gemeinsam die Fehldiagnose, die bei Benutzung aller 9 U-en auftreten würde (Tab.1.5.)
- Ekg-Nr.2: Die Anwendung der mDA erfolgt unnötigerweise, da schon die eDA die zutreffende Krankheit identifiziert.
- Ekg-Nr.20: Zur Erkennung der richtigen Krankheit werden nur 4 der 9 möglichen U-en benötigt.
- Ekg-Nr.36: Alle U-en sind notwendig, um die DDA so weit wie möglich zu differenzieren.
- Ekg-Nr. 47: Auch die Anwendung von miSV u n d mDA vermag die richtige Krankheitsklasse Nr.3 nicht zu erfassen.

An diesen Beispielen ist zu erkennen, daß das miSV - im Vergleich zum NSV unter Benutzung a l l e r U-en - nicht nur eine Ersparnis an U-en erbringt, sondern daß es auch, teilweise unter Zuhilfenahme der mDA, Fehldiagnosen vermeiden kann. Jegliche falsche DA zu verhindern, ist aber auch mit miSV und mDA nicht möglich. So ermöglicht der wechselhafte Verlauf der DF bei Ekg-Nr. 47 an keiner Stelle die Anwendung der SR, so daß hier alle möglichen U-en durchgeführt werden und am Schluß die tatsächliche Krankheit (FT) auch durch die mDA nicht erfaßt wird. Überhaupt fällt in diesem Beispiel auf, daß das Verhalten der a posteriori WK-en in keiner Weise auf das Vorliegen von FT hindeutet.

Wendet man das miSV, beginnend mit den BM Nr.5 und 6, jeweils miteDA oder mDA auf das Kollektiv von 80 Ekgs an, so werden (in Klammern die entsprechenden Ergebnisse des NSV unter Benutzung a l l e r U-en)

- bei Anwendung der eDA 12 (11) Fehldiagnosen getroffen:
 $FDR_B = 12/80$ ($11/80$),
- durch die Anwendung der mDA hiervon 9 (5) korrigiert,
 so daß $FDR_M = 3/80$ ($6/80$) resultiert,
- gleichzeitig durch die mDA 11 (11) Ekgs unnötigerweise echt modifiziert beurteilt, was einen MQ von 1,2 (2,2) ergibt,
- im Durchschnitt 3,2 (0) von 9 möglichen U-en eingespart.

Bemerkt sei, daß von den drei durch die mDA falsch klassifizierten Ekgs tatsächlich zwei einen ähnlich wechsel-

haften Verlauf der DF wie Ekg-Nr. 47 zeigten.¹⁾

Sind in diesem Paragraphen exemplarisch für die Basis-U-en Nr.5 und Nr.6 der Verlauf der DF und die Reihenfolge der U-en bei den verschiedenen SV-en sowie die Auswirkungen des miSV auf einige typische Ekg-Beispiele erläutert worden, so soll in den folgenden Paragraphen gezeigt werden, wie sich die SV-en bei Anwendung auf ein Kollektiv von 80 Patienten und bei Variation des Basis-U-en verhalten. Dabei sollen die Kriterien für die U-Verfahren - FDR_B für die eDA, FDR_M , Q_n , ΔINF und MQ für die mDA - wie in Kap. 2.3. in Abhängigkeit der jeweils durchschnittlich benötigten U-Anzahl diskutiert werden. Wie bei der Behandlung der NSV-en ist es auch hier dazu notwendig, verschiedene Merkmalskombinationen - hier die Basis-U-en (BM) - vorzugeben und die sich ergebenden Werte der genannten Kriterien für die einzelnen SV-en und NSV-en zu vergleichen.

1) Verzichtet man bei der Wahl der SR, wie KNILL-JONES (1973), auf die hier verlangte Stabilität der DF, bekommt man zusätzlich noch die durch " Ausreißer " bedingten Fehldiagnosen (vgl. Tab. 1.4., Ekg Nr. 31). So ist ein großer Teil der FDR bei KNILL-JONES auf solche Fälle zurückzuführen.

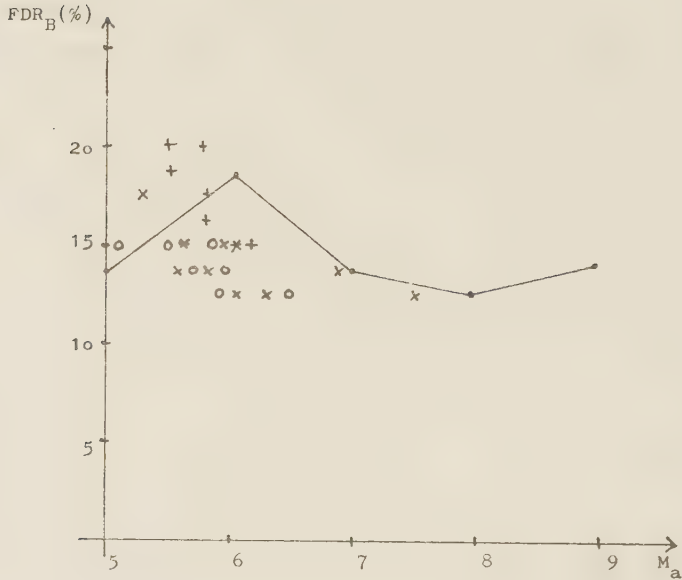


Diagramm 3.2.

Die FDR der eDA (FDR_B) in Abhängigkeit der durchschnittlich benötigten U-Anzahl M_a für das "beste" NSV (.), miSV (o), mdSV (x), uiSV (+) und udSV (*). Verschiedene Punkte stehen für verschiedene Vorgabe der Basis-U-en.

Für gleiches M_a liegt FDR_B für fast alle SV-en (außer für uiSV) niedriger als für das "beste" NSV. Dabei liegt M_a für die SV-en am Beginn des Intervalls, in dem sich FDR_B für das "beste" NSV nur noch wenig ändert ($5 \leq M_a \leq 9$).

3.3.2. eDA: Vergleich der Selektionsverfahren

Schon beim Vergleich von eDA und mDA in Kap. 2.3.1. haben wir gesehen, daß FDR_B bei Anwendung von NSV-en ab einer U-Anzahl von $M_a=5$ in eine stabile Phase eintritt. Ab hier liegt FDR_B zwischen 12,5% und 18,5% (wobei zu berücksichtigen ist, daß durch einen Zufallseffekt der Wert für $M_a=4$ etwas zu niedrig, der für $M_a=5$ etwas zu hoch ist, vgl. Kap. 4.1.). Die Werte von FDR_B sind auch in Diagramm 3.2., in dem die FDR der eDA für die einzelnen SV-en vergleichsweise dargestellt wird, wieder für die NSV-en durch einen Kurvenzug miteinander verbunden.

Bei dieser Darstellung beschränken wir uns auf das Intervall $5 \leq M_a \leq 9$, da in diesen Bereich die durchschnittlich benötigte U-Anzahl aller SV-en fällt, hier wiederum bevorzugt in das Intervall $5 \leq M_a \leq 6,5$. Es fällt auf, daß FDR_B dabei bei Anwendung der multivariaten SV (mi- und mdSV) kaum eine Variation zeigt. Unabhängig von der Vorgabe von BM liegt FDR_B zwischen 12,5% und 15% und - bis auf einen Fall - stets unterhalb der NSV-Kurve. Dasselbe gilt überraschenderweise auch für die Ergebnisse des einfachsten SV, udSV, während FDR_B für uiSV meist größer als die entsprechenden Werte des NSV ist.

Abgesehen von uiSV zeigen also alle SV-en und das NSV kaum eine Variation der FDR der eDA im dargestellten Intervall $5 \leq M_a \leq 9$. Wie schon in Kap. 2.3. erwähnt, liegt der nicht zu unterschätzende Vorteil der SV-en gegenüber dem NSV für die eDA darin , daß ihre U-Anzahl überwiegend am Beginn der stabilen Phase von FDR_B liegt. Hier wird erstmals die minimale FDR erreicht, die auch durch

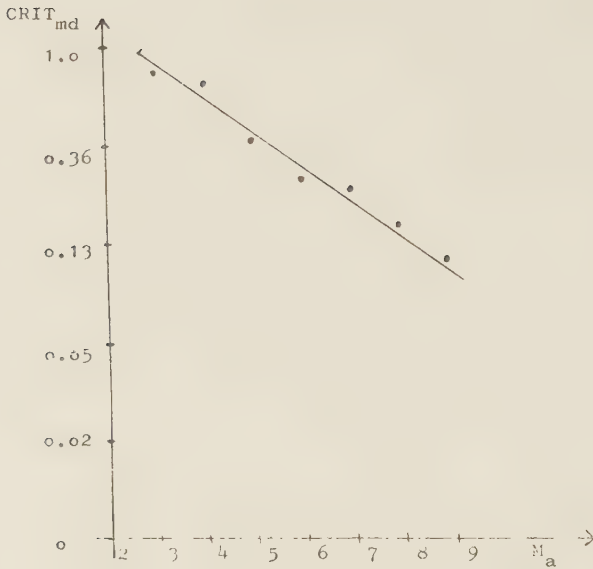


Diagramm 3.3.

$CRIT_{md}$ (AM, m) in Abhängigkeit von der jeweils aktuellen U-Anzahl M_a für das "beste" NSV (bei Basis-U-en Nr.5 und Nr.6) im halblogarithmischen Maßstab.

Das zur Auswahl des "besten" NSV benutzte Kriterium $CRIT_{md}$ läßt die stabile Phase von FDR_B ($5 \leq M_a \leq 9$) nicht vorab erkennen. Es deutet im Gegenteil auf eine monoton fallende FDR der eDA hin.

die Hinzunahme weiterer U-en nicht mehr verringert werden kann. Während diese Grenze also von den SV-en i.a. automatisch erreicht wird, wird sie bei Benutzung eines NSV höchstens einmal zufällig getroffen. Stellt man einen Gütevergleich der verschiedenen Verfahren bzgl. des entscheidenden Kriteriums der eDA, FDR_B , auf, bekommt man nach Diagramm 3.2. folgende Reihenfolge (mit steigendem FDR_B): miSV, mdSV, udSV, NSV, uiSV.

Entgegen aller Erwartung ist auch die Berechnung von $CRIT_{md}$ (Kap. 3.2.2.) für alle Merkmalkombinationen nicht in der Lage, den Beginn der stabilen Phase vorab anzuzeigen. Zwar kann dieser Wert das jeweils durchschnittlich beste Merkmal auf jeder Stufe der DF sehr gut bestimmen, zum Vergleich der FDR_B für v e r s c h i e d e n e M_a eignet er sich aber nicht. Verfolgt man nämlich $CRIT_{md}(AM, m)$ mit wachsendem M_a (in Diagramm 3.3. für $BM=(5,6)$), so sieht man, daß dieser Wert monoton, in halblogarithmischem Maßstab sogar linear, abnimmt: von einem Eintritt in eine stabile Phase ist nichts zu erkennen.

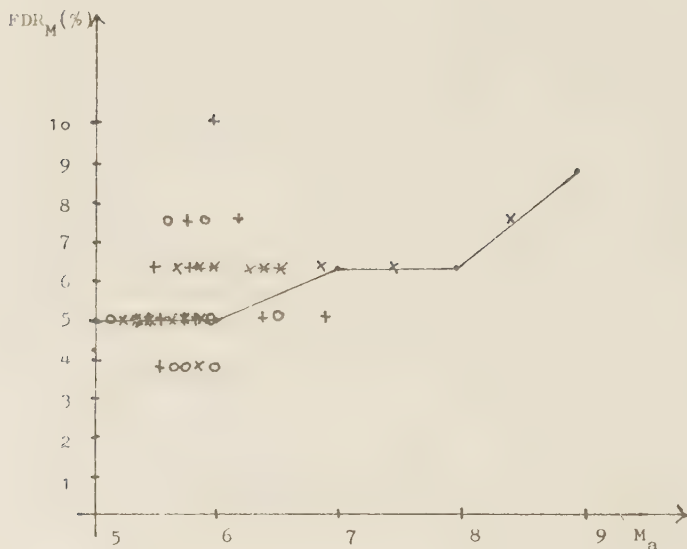


Diagramm 3.4.

FDR_M in Abhängigkeit der U-Anzahl M_a für das "beste" NSV (•), das miSV (○), mdSV (×), uiSV (+) und das udSV (*).

Bei Anwendung der mDA unterscheiden sich die FDR -en im Intervall $5 \leq M_a \leq 9$ nur geringfügig. Das niedrigste FDR_M (3,75%) wird jedoch am häufigsten von dem miSV erreicht - bei Benutzung von nur ca. 6 von maximal 9 U-en.

3.3.3. mDA: Vergleich der Selektionsverfahren

Betrachtet man die FDR der mDA, FDR_M für die SV-en in analoger Weise (Diagramm 3.4.), stellt man folgendes fest: Im Vergleich liegen die FDR-en des miSV zum größeren Teil unterhalb denen des NSV ("gute" BM), zum geringeren Teil oberhalb ("schlechte" BM), die des mdSV entsprechen denen des NSV ungefähr, liegen eher etwas ungünstiger. Die Anwendung der univariaten SV-en resultiert wieder in höheren FDR-en, wobei jedoch das uiSV vergleichsweise besser abschneidet als bei Benutzung der eDA. Insgesamt ist auch hier der Unterschied nur gering: fast alle Werte liegen zwischen 3,75% und 7,5%. Eine Reihenfolge in der Güte der verschiedenen Verfahren kann (mit wachsenden FDR_M) so angegeben werden: miSV ("gute" BM), NSV, mdSV, udSV, uiSV, miSV ("schlechte" BM).

Bei Anwendung der mDA ist die FDR jedoch nur ein Gütekriterium, das andere ebenso wichtige ist die Rate derjenigen Ekgs, die durch die mDA einer echt mDA zugeführt werden (Q_n , Diagramm 3.5.) bzw. der Verlust an diagnostischer Schärfe, den man durch den teilweisen Verzicht auf eine weitergehende Differenzierung erleidet. (ΔINF , Diagramm 3.6.). Beide Maße ergeben für die Gütebeurteilung der einzelnen Verfahren identische Ergebnisse: Waren die Vorteile der SV-en gegenüber dem NSV bei der Betrachtung der FDR-en kaum feststellbar, so kommt ihr Nutzen nun deutlich zum Ausdruck. Die diagnostischen Aussagen verlieren bei Anwendung jedweden SV erheblich weniger an DS als bei Benutzung eines NSV. Dies gilt in

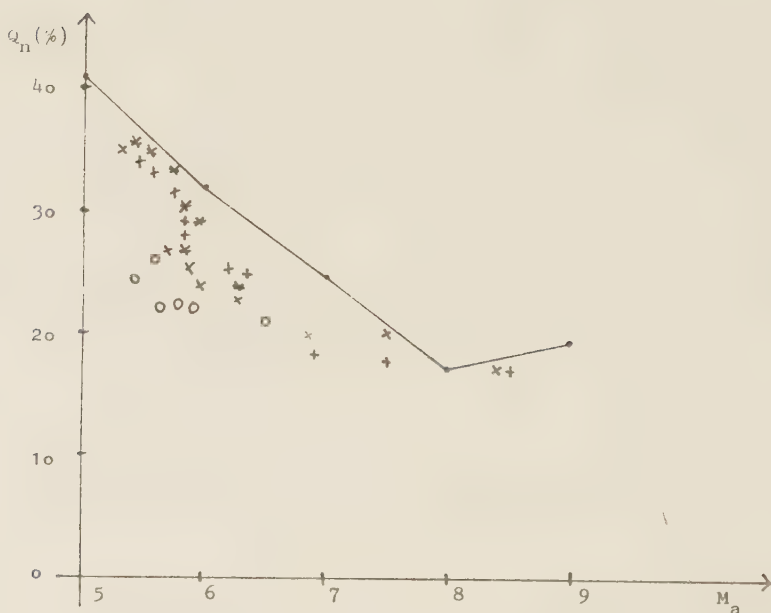


Diagramm 3.5.

Rate der echt mDA-en (Q_n) in Abhängigkeit der durchschnittlich benutzten U-Anzahl M_a für das "beste" NSV (•), das misV (○), mdSV (×), uiSV (+) und udSV (*).

Zur Reduktion von FDR_B auf jeweils ähnliche Werte von FDR_M (vgl. Diagramm 3.4.) benötigen die SV-en ausnahmslos weniger echt mDA-en als das "beste" NSV.

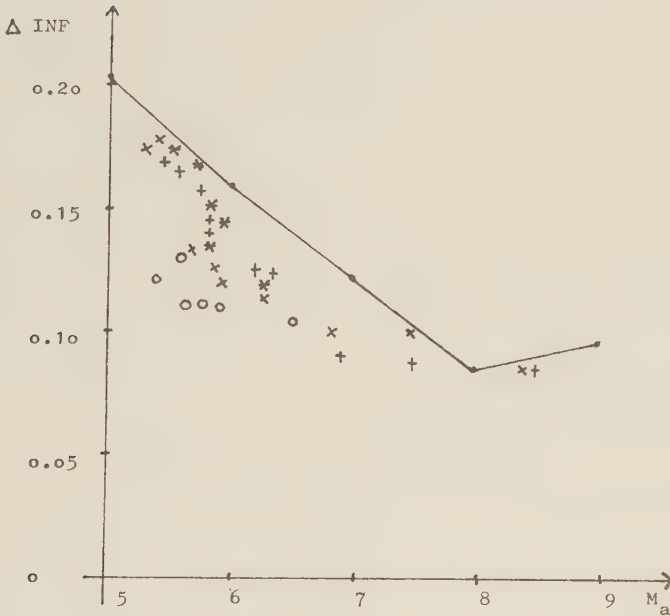


Diagramm 3.6.

ΔINF in Abhängigkeit von M_a für das "beste" NSV ($\cdot$), das miSV ($\circ$), mdSV ($\ast$), uiSV ($+$) und udSV ($\times$). Wie bei Q_n (Diagramm 3.5.) zeigt der VaDS auch hier für alle SV-en niedrigere Werte als für das "beste" NSV. Auffallend ist die Gruppierung der Punkte für die individuellen SV-en an der unteren Grenze des Punktehaufens.

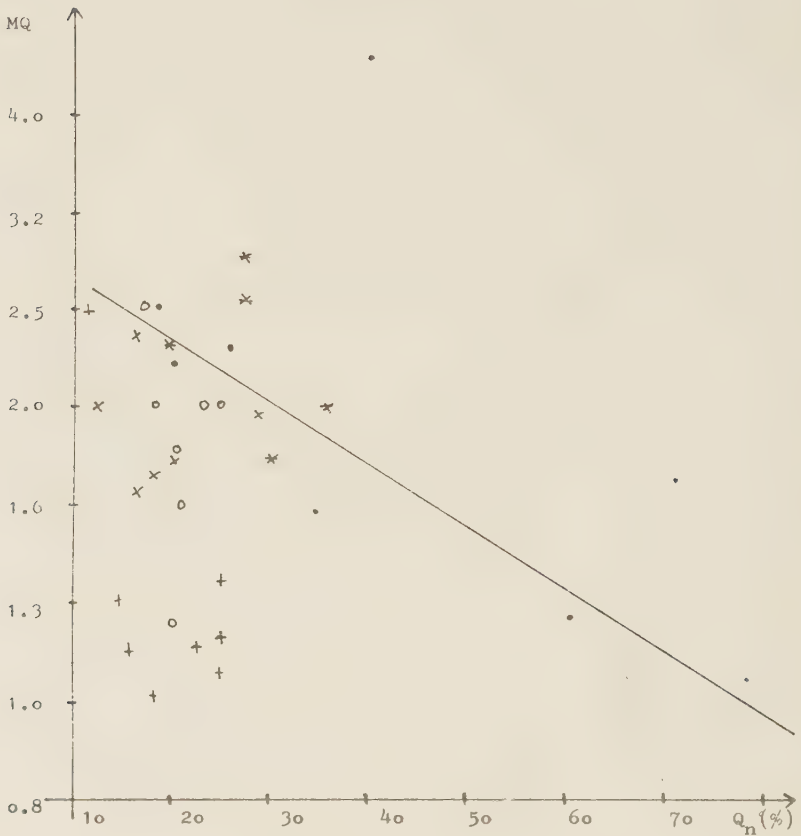


Diagramm 3.7.

Modifikationsquotient MQ in Abhängigkeit der Rate echt mDA-en (Q_n) für das "beste" NSV (•, gemittelt durch Kurvenzug), miSV (○), mdSV (x), uiSV (+), udSV (*) in halblogarithmischer Darstellung.

Die Anwendung der mDA erbringt für fast alle SV-en (besonders das uiSV) in der Bilanz des MQ günstigere Ergebnisse als für das "beste" NSV.

besonderem Maße für das miSV, das seine Überlegenheit klar demonstriert. Dies liegt daran, daß es (vgl. Kap. 3.3.1.) am ehesten eine Verstärkung der DF gewährleistet und diese nur dann stoppt, wenn eine weitere Differenzierung nicht mehr sinnvoll ist. Einen außerordentlich geringen VaDS bekommt man dabei bei Anwendung der "schlechten" BM, die somit ihren Nachteil bei der FDR der mDA teilweise kompensieren. Die weitere Reihenfolge ergibt sich dann (mit wachsendem VaDS) wie folgt: miSV ("schlechte" BM), miSV ("gute" BM), mdSV, uiSV, udSV, NSV.

Schließlich zeigt Diagramm 3.7. mit der Darstellung des Modifikationsquotienten die schon in Kap. 2.3.3 erwähnte Tatsache, daß sich die Anwendung der mDA für individuelle SV besonders eignet, mit großer Deutlichkeit. Besonders miSV und uiSV profitieren von der Möglichkeit, sich eine eindeutige DA in bestimmten Fällen vorzubehalten. Bei dem uiSV - einem Verfahren, das, wie später (Kap. 3.4 und 4.4.) noch ausgeführt wird, mit ärztlichem Diagnoseverhalten große Ähnlichkeit hat - müssen meist nur 1 bis 1,3 Fälle unnötigerweise reserviert beurteilt werden, um eine Fehldiagnose der eDA zu korrigieren. Auch das mdSV schneidet noch günstiger ab als das NSV, während das udSV durch die mDA keinen großen Vorteil erfährt. Insgesamt erhält man (mit wachsendem MQ) folgende Ordnung: uiSV, miSV, mdSV, NSV, udSV.

3.4. Schlußfolgerungen

Die bei der DD kongenitaler Hervitien erzielten Ergebnisse für die Anwendung der konstruierten SV-en kann man abschließend wie folgt beurteilen: Bei Benutzung der eDA besteht der Vorteil der SV-en gegenüber dem NSV im wesentlichen darin, daß die "optimale" U-Auswahl und U-Anzahl i.a. automatisch erreicht wird. Die FDR dieser Verfahren bietet, verglichen mit der des NSV, hingegen kaum eine Verbesserung, beim uiSV sogar eine deutliche Verschlechterung. Da dieses uiSV - Vergleich der einzelnen U-en (bei jedem Schritt der DF) getrennt voneinander (univariat) und unter Berücksichtigung des individuellen Falles - dem ärztlichen Diagnoseverhalten sehr ähnlich ist (vgl. Kap. 4.4.), entspricht dies der Erfahrung , die STERN et al. (1974) machten und die schon zu Beginn von Kap. 2 erwähnt wurde: Bei einem solchen Diagnoseverhalten bewährt sich die Anwendung der mDA in besonderer Weise, so bekommen wir analog zu STERNs Beobachtungen einen äußerst günstigen MQ beim uiSV.

Allerdings reicht auch diese Verbesserung für das uiSV nicht aus, um (bei Benutzung der mDA) auch nur annähernd so gute Ergebnisse wie die multivariaten SV-en zu erreichen. Hier zeichnet sich besonders das miSV dadurch aus, daß es bei minimaler FDR das größtmögliche Differenzierungsvermögen erreicht und die geringste Anzahl an U-en dafür benötigt. Daß der Gewinn dabei hauptsächlich in der DS erfolgt, ist nicht zwangsläufig: Über eine Variation der Klassifikationskonstanten c_j (nach Gl. (2-1)) kann man

nämlich die mDA so ändern, daß der VaDS größer - etwa in der Größenordnung mit den anderen SV-en vergleichbar - wird, FDR_M aber weiter sinkt.

Bei Benutzung der mDA stellt auch das mdSV noch eine brauchbare Alternative zum miSV dar, ohne dessen Nutzen ganz zu erreichen. Es bietet aber den Vorteil, einen geringeren Rechenaufwand zu fordern und ist deshalb möglicherweise bei sehr großer U-Anzahl angezeigt. Ähnliches gilt für das udSV, das aber wegen des ungünstigen MQ in erster Linie bei Anwendung der eDA in Betracht kommen sollte. Insgesamt sieht man, daß das eingeführte miSV die Forderungen, die in den Beispielen von Kap. 1.3. und zu Beginn von Kap. 3 an ein gutes Diagnoseverfahren gestellt wurden, in weitem Umfang erfüllt, so daß es offenbar - besonders in Verbindung mit der mDA - eine nützliche und praktikable Verbesserung der bisher gebräuchlichen Computer-gestützten Klassifizierungsmethoden beinhaltet.

KAPITEL IV

WEITERE EIGENSCHAFTEN UND VARIATIONSMÖGLICHKEITEN DES NEUEN VERFAHRENS

4.1. Einfluß der Basisuntersuchungen

4.1.1. Einleitung

Der Einfluß der Basis-U-en (oder Basismerkmale BM) auf die Kriterien der DF ist in den bisherigen Kapiteln häufig erwähnt, zugunsten anderer Fragestellungen aber bis hierhin zurückgestellt worden. Das Problem, welche und wieviele U-en man in BM routinemäßig durchführen sollte, ist auch für die äDF immer wieder diskutiert worden. So ist eine geringe Anzahl von U-Ergebnissen zwar leichter übersehbar, eine hierauf aufbauende U-Abfolge kann aber leicht zur Fehldiagnose führen. Andererseits ist (GROSS, 1966) ein umfangreicher "check-up" für den Arzt nicht selten unübersichtlich und verwirrend. Wie sich diese Fragestellung bei der CDF darstellt, soll im folgenden erläutert werden.

In Kap. 3 wurden die SV-en für verschiedene BM stets mit den Ergebnissen der "besten" NSV-en verglichen, die sich mit Hilfe des Kriteriums $CRIT_{md}$ (Kap. 3.2.2.) bestimmen ließen. Dabei bekam man, für eine vorgegebene Merkmalsanzahl M_a , die für dieses M_a beste Kombination von U-en, indem man die ersten M_a Elemente der durch $CRIT_{md}$ ermittelten Merkmalsfolge 6,2,10,1,3,5,4,9,11 betrachtete. Die für diese "besten" NSV-en erhobenen Resultate sollen auch in den Diagrammen dieses Paragraphen wieder durch einen Kurvenzug miteinander verbunden sein.

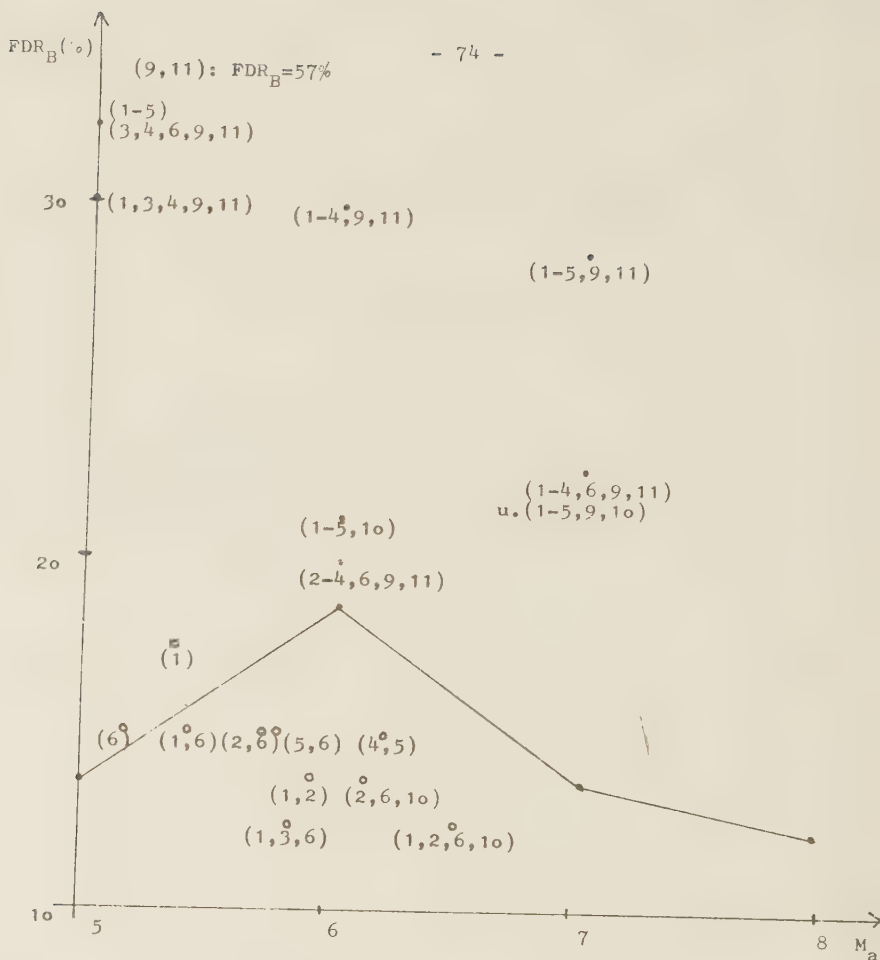


Diagramm 4.1.

FDR_B bei verschiedener Vorgabe von AM (NSV, $\cdot$) und BM (miSV, $\circ$).
in Abhängigkeit der U-Anzahl. Die in Klammern angegebenen Zahlen
bezeichnen die jeweils vorgegebenen U-Kombinationen.
Alle Punkte der übrigen NSV-en liegen höher als die der "besten"
NSV-en (Kurvenzug) und diese wiederum höher als fast alle
Werte der miSV-en. Im Gegensatz zu den NSV-en ist die FDR der
EDA für das miSV weitgehend unabhängig von der Vorgabe der U-en.

Daneben sind die Ergebnisse anderer NSV-en wiedergegeben, wobei die jeweils benutzte Merkmalskombination unter den entsprechenden Punkten in Klammern angegeben ist. Da es eine große Anzahl solcher Kombinationen gibt (für jedes M_a genau $\binom{M}{M_a}$), ist in den Diagrammen nur eine Auswahl von insgesamt 10 NSV-en aufgeführt.

Dasselbe gilt für die Auswahl der BM bei der Diskussion des miSV - nur dieses SV soll hier betrachtet werden. Hier geben die Zahlen in Klammern an, von welchen BM ausgehend das SV durchgeführt wurde. Die zugehörige U-Anzahl M_a stellt dann die im Durchschnitt der 80 betrachteten Ekgs benötigte Anzahl von U-en dar. Zur Darstellung kommen soll in diesem Paragraphen wieder nur das Intervall $5 \leq M_a \leq 8$, da in diesem Bereich die Ergebnisse aller relevanten SV-en liegen.

4.1.2 eDA: Variation von BM

Daß die "besten" NSV-en diese Bezeichnung auch tatsächlich verdienen, zeigen die in Diagramm 4.1. eingezeichneten Werte für FDR_B . Ihre Ergebnisse liegen deutlich besser als die der anderen NSV-en, teilweise sind ihre FDR-en nur halb so groß. Außerdem fällt auf, daß das Auswechseln nur eines Merkmals (etwa bei $M_a=6$: Nr.6 statt Nr.1) erhebliche Auswirkungen haben kann. Die Hinzunahme einer U kann, wie die von Nr.10 zur Kombination (1-5), die FDR stark verringern, sie kann sie aber auch, wie beim Übergang von $M_a=6$ zu $M_a=7$ in zwei Fällen, vergrößern. Dagegen ist ein solch wechselhaftes Verhalten bei Anwendung von miSV nicht festzustellen. FDR_B ist hier weitgehend

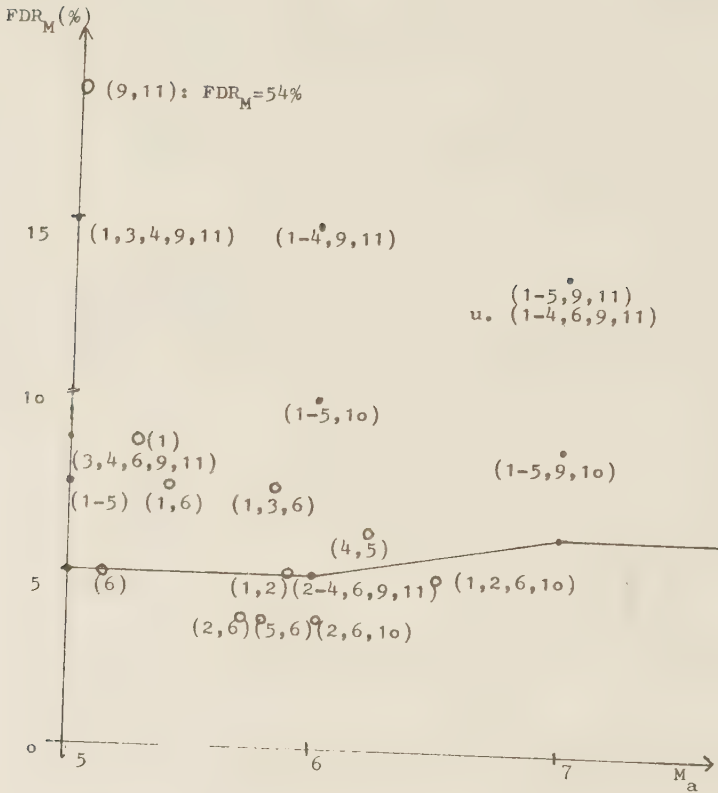


Diagramm 4.2.

FDR_M in Abhängigkeit von M_a (Bezeichnungen wie in Diagramm 4.1.)

Bei Anwendung des NSV zeigt FDR_M wie FDR_B eine starke Abhängigkeit von der Vorgabe der U-en, während es für das miSV hiervon weitgehend unabhängig ist.

unabhängig von der Wahl von BM und liegt meist zwischen 12,5% und 15%. Nur die Kombination (9,11) - nach $CRIT_{md}$ die beiden schlechtesten U-en - macht eine Ausnahme; ihre FDR liegt mit 57% extrem hoch.

4.1.3. mDA: Einfluß von BM

Bei Anwendung der mDA zeigen sich dieselben Erscheinungen wie bei der eDA: Die Ergebnisse der NSV-en sind in allen Kriterien, FDR_M und VaDS, stark schwankend bei Änderung der Merkmalskombinationen, das miSV dagegen sehr stabil gegen Variation von BM. Bei der Darstellung von FDR_M (Diagramm 4.2.) wiederholen sich im wesentlichen die entscheidenden Phänomene, die schon bei Betrachtung von FDR_B auftraten, auf einem niedrigeren Niveau. Hier fällt lediglich auf, daß die FDR des miSV für $BM=(1,6)$ und $BM=(1,3,6)$ ("schlechte" BM aus Kap. 3.3.3.) über der Kurve der "besten" NSV-en zu liegen kommt. Dafür erweisen sich diese BM bei der Diskussion des VaDS als besonders günstig, bei ihrer Benutzung ist sowohl Q_n (Diagramm 4.3.) als auch ΔINF (Diagramm 4.4.) sehr niedrig. Überhaupt zeigt sich in diesen beiden Diagrammen wieder, daß die Anwendung der mDA beim miSV nur einen kleinen VaDS erbringt. Auffallend ist dabei, wie gleichmäßig alle Werte für ΔINF fast auf einer horizontalen Linie liegen, obwohl einige BM Merkmale enthalten, die in der durch $CRIT_{md}$ bestimmten Hierarchie nicht an vorderer Stelle liegen, wie etwa $BM=(4,5)$. Nur $BM=(9,11)$ folgt dieser Regelmäßigkeit nicht; ihrem äußerst niedrigen

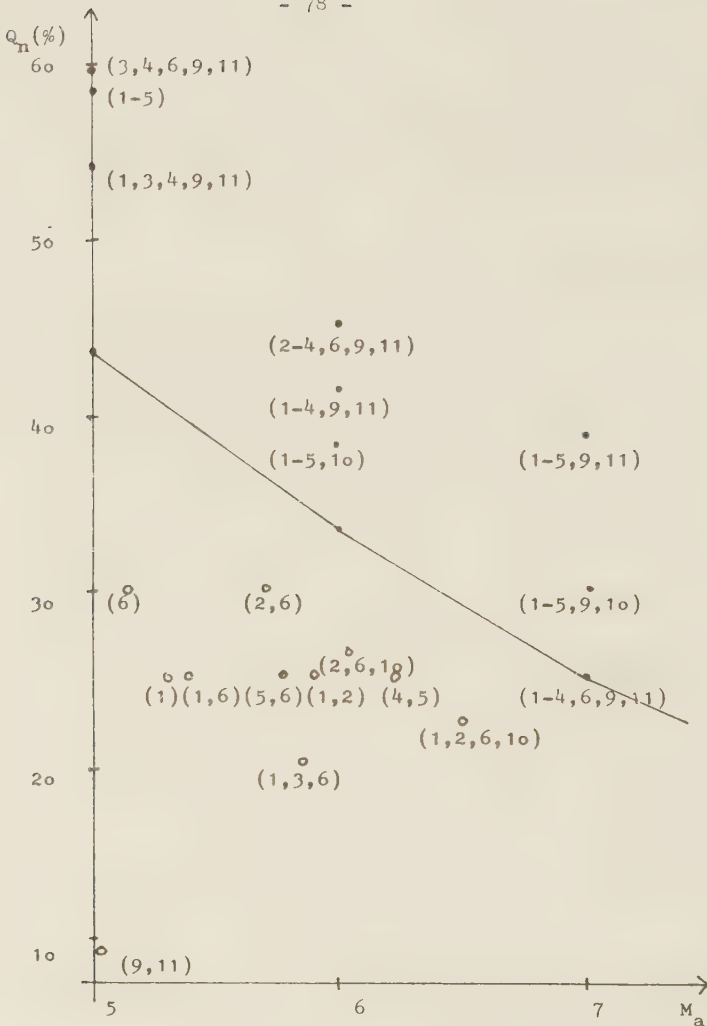


Diagramm 4.3.

Q_n in Abhängigkeit von M_a (Bezeichnungen wie in Diagramm 4.1.)
 Wieder sind die Werte für das miSV weitgehend unabhängig von der Wahl von BM. Q_n liegt für das miSV deutlich niedriger als für das NSV, wobei die Kurve der "besten" NSV-en eine scharfe Grenze zwischen den Werten der übrigen NSV-en und denen der miSV-en zieht.

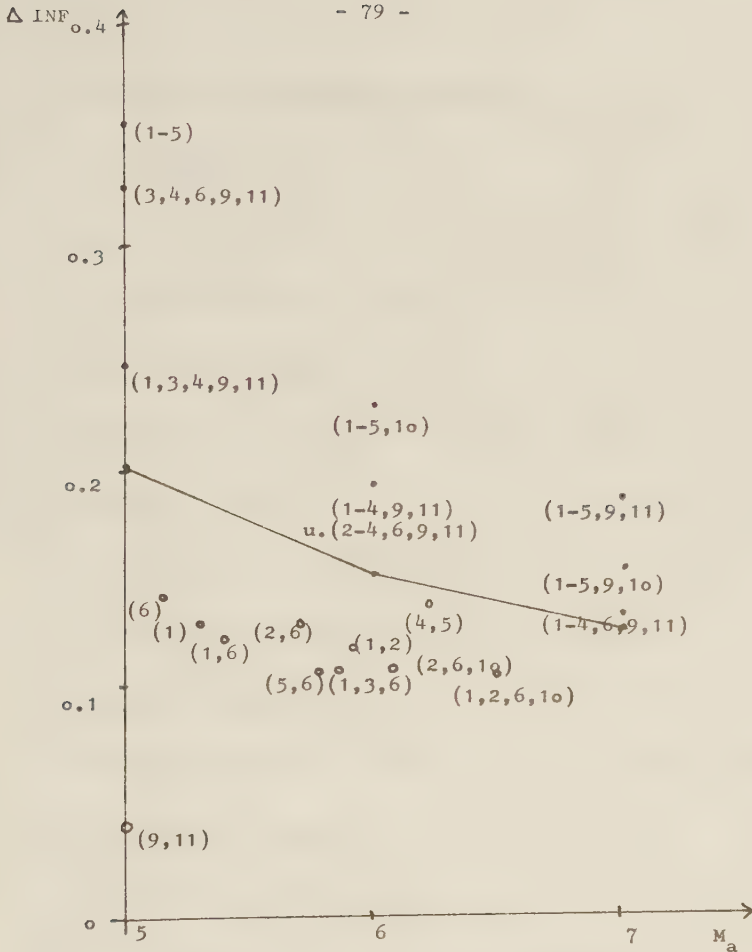


Diagramm 4.4.

ΔINF in Abhängigkeit von M_a (Bezeichnungen wie in Diagramm 4.1.)

Wie in den Diagrammen 4.1. - 4.3. zeigt auch INF bei Anwendung des miSV weitgehende Unabhängigkeit von der Wahl der Basis-U-en: Fast alle Werte für den VaDS liegen für das miSV im Intervall $0.1 \leq \Delta \text{INF} \leq 0.15$.

VaDS steht die riesige FDR von 54% gegenüber, was diese BM für die Klassifikation unbrauchbar macht.

Zeigen sich auch alle anderen BM bei Anwendung des miSV annähernd gleichwertig in allen Kriterien sowohl der eDA als auch der mDA, so scheint doch die geringere Anzahl benötigter U-en für den Beginn mit einer Kombination von zwei U-en zu sprechen. Dafür spricht auch die Tatsache, daß man mit 2 Basismerkmalen gerade eine erste "echte" DDA bekommt, falls man die initialen U-en so auswählt, daß sie alle vier Klassen gleichmäßig trennen - was man wiederum über die univariaten BK-en (vgl. Kap. 3.2.1.) bewerkstelligen kann. Ein Beginn mit nur einem BM dagegen birgt stets die Gefahr einer sehr "zufälligen" initialen DDA, der mit drei oder mehr BM eine frühe Festlegung auf eine bestimmte Konstellation von Krankheiten in sich. Wenn diese Überlegungen auch durch die hier angeführten Ergebnisse nicht eindeutig zu belegen sind und an einer Fragestellung mit mehr als vier differentialdiagnostischen Möglichkeiten zu untersuchen sind, so zeigen sich doch in den hier demonstrierten Resultaten die Vorteile unseres neuen Verfahrens, besonders eindrucksvoll bei der Betrachtung der Diagramme 4.3. und 4.4.: Die Kurve der "besten" NSV trennt klar die Punkte der übrigen NSV-en von denen des miSV.

4.2. Wichtung von Krankheitsklassen

4.2.1. Einleitung

Bis hierhin sind wir davon ausgegangen, daß alle in der initialen DDA vorgegebenen Erkrankungen gleichwertig zu behandeln sind, daß ihnen jeweils die gleiche Bedeutung beizumessen ist. Häufig ist man aber bemüht, bestimmte Krankheiten auf keinen Fall zu übersehen, während man auf das Erkennen anderer geringeren Wert legt. So steht bei ALPEROVITCH (1977), der dieses Problem für drei verschiedene Situationen erläutert hat, etwa bei der DD eines Rundherdes in der Lunge die Frage nach einem Bronchial-Carcinom an vorderster Stelle. Will man das Problem der Wichtung bestimmter K-en in das hier vorgestellte miSV einführen, werden zunächst einige theoretische Überlegungen notwendig.

4.2.2 Theoretische Grundlagen

Diese Wichtung erfolgt in statistischen Entscheidungsverfahren mit Hilfe einer Verlustfunktion L , indem man den Verlust, den man bei Nichterkennen der Krankheit Nr. j erleidet, mit L_j quantifiziert. HESS (1979 A) hat gezeigt, daß dieses Problem identisch ist mit der Veränderung der tatsächlichen relativen Häufigkeiten p_j zu "künstlichen" p_j' . Diese p_j' sind, verglichen mit p_j , umso größer, je wichtiger das Erkennen der K Nr. j ist.

Genau lassen sie sich nach der Formel

$$p_j' = p_j L_j / \sum_{k=1}^N L_k \quad \text{Gl. (4-1)}$$

berechnen. In unseren Beispielen wichten wir jeweils eine K Nr.i mit $L_i=10$ und die jeweils übrigen mit $L_j=1$ (für $j \neq i$). Für die Klassifikation eines gemessenen Ekgs x sind dann die "künstlichen" a posteriori WK-en

$$q' (j | x^{(AM)}) = p_j' f(x^{(AM)} | j) / (\sum_{k=1}^N p_k' f(x^{(AM)} | k)) \quad \text{Gl. (4-2)},$$

falls die aktuellen Merkmale AM für x berechnet worden sind. Diese Änderung von q zu q' hat für die Durchführung des miSV drei Konsequenzen:

(A) Bei Anwendung der eDA führt die Vergrößerung von $q (i | x^{(AM)})$ zu $q' (i | x^{(AM)})$ für $L_i=10$ und $L_j=1$ (für $j \neq i$) gemäß Gl. (1-7) zu einer bevorzugten Zuordnung in K Nr.i. Diese Krankheit wird nun häufiger als in der Situation ohne Wichtung diagnostiziert, die anderen seltener. Gleichzeitig sinkt die WK, K Nr.i nicht zu erkennen ($FDR_{B,G} (i)$) gegenüber $FDR_B(i)$ (nach Gl.(1-5)), während für $j \neq i$ $FDR_{B,G}(j)$ größer als $FDR_B(j)$ wird. Da die gewichtete eDA keine "Bayes" Entscheidung mehr darstellt, wird auch die durchschnittlich zu erwartende $FDR_{B,G}$ größer als FDR_B (nach Gl.(1-8)).

(B) Die Anwendung von Gl.(2-1) der mDA mit geänderten a posteriori WK-en führt wie in (A) zu einer bevorzugten Klassifizierung in K Nr.i (falls $L_i=10$). Dabei ist die Bedingung $q' (i | x^{(AM)}) \gg c_i$ äquivalent zu der Forderung $q (i | x^{(AM)}) \gg c_i'$:
$$= \frac{c_i}{L_i} \frac{1}{1 - c_i \frac{L_i-1}{L_i}}$$

Gl. (4-3)

Für große L_i und kleine c_i ($L_i \gg 1$, $c_i \ll 1$) wird c_i' dann annäherungsweise zu $c_i' \sim c_i/L_i$ (in unserem Beispiel also $c_i \sim 0.01$). Eine ähnliche Beziehung wie in Gl. (4-3) läßt sich für $j \neq i$ (d.h. $L_j=1$) nicht formulieren, so daß die gewichtete mDA keine mDA mit veränderten Klassifikationskonstanten darstellt. An der Verringerung von $q'(j|x^{(AM)})$ gegenüber $q(j|x^{(AM)})$ für $j \neq i$ erkennt man aber, daß eine Zuordnung in K Nr. j erschwert wird.

(C) Da bei der U-Auswahl des miSV (Punkt (9) in Diagramm 3.1.) die aktuellen a posteriori WK-en entscheidend eingehen, verändert die Variation von $q_A(j)$ zu $q_A'(j)$ auch die Selektion der nächstbesten U. Nun werden die U-en bevorzugt benutzt, die K Nr. i am besten von den anderen K-en differenzieren.

4.2.3. Ergebnisse

Wir haben, wie erwähnt, bei der DD kongenitaler Herzvitien nach einander je eine K Nr. i mit $L_i=10$, die jeweils anderen mit $L_j=1$ (für $j \neq i$) gewichtet. Die Veränderung der tatsächlichen a priori WK-en zu den "künstlichen" p_j' (nach Gl. (4-1)) zeigt Tab. 4.1.:

Tab. 4.1.

gewichtete K Nr. i ($L_i=10$)	p_i	"künstliche" a priori WK-en			
		p_1'	p_2'	p_3'	p_4'
1	0.21	0.73	0.10	0.07	0.10
2	0.29	0.06	0.79	0.05	0.09
3	0.19	0.08	0.11	0.70	0.11
4	0.31	0.06	0.08	0.05	0.81

"Künstliche" a priori WK-en bei Wichtung von K Nr. i
Durch Wichtung von K Nr. i wird diese K in einer Häufigkeit angenommen (p_i'), die die tatsächliche (p_i) weit übersteigt.

Tab. 4.2.

gewichtete K Nr.i ($L_1=10$)	gewichtete eDA		ungewichtete eDA	
	$FDR_{B,G}(i)$	$FDR_{B,G}$	$FDR_B(i)$	FDR_B
1	0	19	5	12
2	0	16	1	12
3	1	14	2	12
4	1	14	4	12

Vergleich der FDR-en von gewichteter und ungewichteter eDA: Während die Patienten, die tatsächlich die jeweils gewichtete K Nr.i haben, durch die gewichtete eDA fast alle der richtigen Diagnose zugeführt werden, sind die entsprechenden Größen der ungewichteten eDA erheblich größer ($FDR_{B,G}(i) < FDR_B(i)$). Dagegen übersteigt die durchschnittliche FDR der gewichteten eDA die der ungewichteten eDA beträchtlich ($FDR_{B,G} > FDR_B$).

Tab. 4.3.

		Zugrunde liegende K Nr.i			
		1	2	3	4
Ekg-Zuordnung in K Nr.j	1	17	4	9	6
	2	0	19	0	0
	3	0	0	6	0
	4	0	0	0	19

Klassifikationsmatrix bei Wichtung von K Nr.1 mit $L_1=10$

(Anwendung der eDA): Alle falsch zugeordneten Ekg's werden der K Nr.1 zugeordnet. Andererseits werden alle Patienten mit K Nr.1 durch die gewichtete eDA der richtigen Diagnose zugeführt.

Man sieht, daß die jeweils mit $L_1=10$ gewichtete K Nr.i in einer Häufigkeit angenommen wird, die die tatsächliche erheblich übersteigt, und die übrigen p_j stark reduziert werden. Wendet man diese künstlichen p_j' auf das miSV (bei Vorgabe der Basis-U-en Nr. 5,6) an, bekommt man folgende Resultate:

(A) Benutzt man die eDA in ihrer nach Kap.4.2.2. gewichteten Form, bekommt man die in Tab. 4.2 aufgeführten FDR-en. Bei Wichtung von K Nr.i mit $L_1=10$ wird für $i=1,2$ die entsprechende K stets, für $i=3,4$ bis auf eine falsche Zuordnung richtig erkannt ($FDR_{B,G}(i)=0$ oder $1/80$). Im Vergleich zu den entsprechenden Größen , die man ohne Wichtung erhält ($FDR_B(i)$), erfüllt bei Anwendung der eDA die angewendete Methode der Wichtung ihre Funktion voll. Fast alle Patienten, die tatsächlich die schwerpunktmäßig behandelte K haben, werden der richtigen Diagnose zugeführt. Umgekehrt ordnet man eine größere Zahl von Patienten, die diese Krankheit nicht haben, K Nr.i zu, was sich in einer Vergrößerung der durchschnittlichen FDR_B zu $FDR_{B,G}$ äußert. Daß hierfür, wie in den theoretischen Überlegungen erwartet wurde, wirklich eine bevorzugte Zuordnung in K Nr.i verantwortlich ist, zeigt die Klassifikationsmatrix in Tab. 1.3.(für den Fall $L_1=10$). A l l e falsch diagnostizierten Ekgs werden in K Nr.1 klassifiziert, von den Patienten, die K Nr.3 haben, sogar 60%. Andererseits werden alle Ekgs von Patienten mit ASD (K Nr.1) richtig zugeordnet.

(B) Benutzt man die mDA in ihrer nach Gl. (4-2) gewichteten Fassung, stellt sich der gesuchte Vorteil gegenüber der ungewichteten mDA jedoch nicht ein. In die jeweils mit $L_i=10$ gewichtete K Nr.i wird zwar wegen kleiner Klassifikationskonstanten c_i' die Zuordnung erleichtert, wesentliche praktische Auswirkungen für das Erkennen dieser K bleiben aber aus. (Tab. 4.4.). Die ungewichtete mDA ist fast genauso wie die gewichtete mDA in der Lage, die Ekgs, die K Nr. i zuzuordnen sind, zu erkennen:

$FDR_{M,G}(i)$ ist nur für $i=4$ geringfügig niedriger als $FDR_M(i)$, ansonsten stimmen beide Größen überein.

Andererseits übersteigen die durchschnittlichen $FDR_{M,G}$ FDR_M beträchtlich, meist ist auch die Anzahl echt mDA-en ($Q_{n,G}$) bei Wichtung größer als ohne zusätzliche Schwerpunktbildung (Q_n). Auch macht die mDA in unserem Beispiel nicht nur eine gewichtete mDA überflüssig, sondern ersetzt auch die gewichtete eDA in vollem Umfang. Der Grund hierfür ist darin zu sehen, daß die Forderung $q(k|x^{(AM)}) > c_k$ eine sehr niedrige Schwelle für das Erkennen der K Nr.k darstellt. Besonders eindrucksvoll zeigt sich dieses Resultat bei Wichtung von K Nr.1 mit $L_1=10$: Sowohl die ungewichtete mDA als auch die gewichtete eDA erkennen alle Ekgs, die K Nr.1 zuzuordnen sind, richtig. Diese bewirkt aber insgesamt 19 Fehldiagnosen, jene nur drei! Sogar die Summe aus FDR_M und Q_n übersteigt $FDR_{B,G}$ nur geringfügig.

Tab. 4.4.

gewichtete K Nr.i ($L_i=10$)	c_i	gewichtete mDA			ungewichtete mDA		
		$FDR_{M,G}(i)$ (n/80)	$FDR_{M,G}$ (n/80)	$Q_{n,G}$ (n/80)	$FDR_M(i)$ (n/80)	FDR_M (n/80)	Q_n (n/80)
1	0.01	0	7	25	0	3	20
2	0.01	0	10	22	0	3	20
3	0.01	1	7	14	1	3	20
4	0.01	1	11	20	2	3	20

Vergleich von gewichteter und ungewichteter mDA: Die Wichtung von K Nr.i erbringt nur für $i=4$ eine geringfügig kleinere Rate von falsch zugeordneten Ekgs der Patienten mit K Nr.i. Dagegen ist die durchschnittliche Rate an Fehldiagnosen ($FDR_{M,G}$) stets und die Rate der echt mDA-en ($Q_{n,G}$) meist bei Anwendung der gewichteten mDA größer als bei der ungewichteten mDA.

(C) Erstaunlich ist bei Durchführung des miSV auch, daß eine Wichtung bestimmter Krankheiten wohl eine veränderte Reihenfolge der U-en bedingt, die durchschnittlich benötigte U-Anzahl aber kaum verändert, wieder werden im Durchschnitt etwa 5 bis 6 der maximal 9 möglichen U-en benötigt.

Das miSV mit mDA eignet sich jedoch nicht nur gut bei verschiedenartiger Wichtung der K-en, sondern auch bei tatsächlicher Änderung der a priori WK-en, wie sie z.B. bei der Übertragung einer CDF in eine andere geographische Region zu erwarten sind. Es ist bekannt, daß sich die FDR-en dabei außerordentlich abhängig von den p_j der einzelnen K-en Nr.j zeigen (PÖPPL, 1977, de DOMBAL, 1975). Für dieses Problem, das auch von

CORNFIELD (1973) und PIPBERGER (1973) für die Ekg-DD von Herzkrankheiten aufgezeigt worden ist, ist von HESS und BRODDA (1981) eine Methode entwickelt worden, die es erlaubt, die Leistungsfähigkeit eines Diagnosesystems v o r seiner Anwendung in einer anderen Region vorherzubestimmen.

4.3. Andere Variationsmöglichkeiten

Die Diskussion der Einflüsse der Basis-U-en und der Wichtung von Krankheitsklassen auf das miSV hat bereits zwei Variationsmöglichkeiten unserer neuen Methode aufgezeigt. Daß diese tatsächlich ein umfassendes und auf die unterschiedlichsten Situationen flexibel anwendbares Verfahren darstellt, sollen die folgenden Überlegungen demonstrieren.

4.3.1. Ausschluß von Untersuchungen

Bei manchen Patienten kann es hilfreich oder sogar notwendig sein, gewisse U-en nicht durchzuführen oder ihre Ergebnisse nicht zu berücksichtigen. Das erste ist der Fall, falls eine U für den Patienten eine zu große Belästigung oder gar Gefährdung darstellt (vgl. auch 4.3.4.), das zweite, wenn eine U, aufgrund systematischer Fehler, ein offensichtlich falsches Ergebnis erbringt. Dann ist es meist günstiger, auf solche U-Ergebnisse zu verzichten.

Dies gelingt bei Anwendung des miSV, indem man bei der Merkmalsauswahl (Punkt (9) in Diagramm 3.1.) eine aufgrund der aktuellen DDA (repräsentiert durch die aktuelle mDA) als gefährlich erachtete $U_{m'}$ nicht berücksichtigt. Offensichtlich falsche U-Ergebnisse läßt man aus, indem man nach Berechnung von $x^{(AM)}$ (Punkt (11) in Diagramm 3.1.) M_a wieder um 1 reduziert und die Merkmalsauswahl mit einem um das entsprechende Merkmal verkleinerten Pool PM wiederholt. Eine solche Reduzierung von PM ist vorab notwendig, wenn man das miSV von einer Klinik auf eine andere übertragen will, in der

gewisse U-en aus technischen Gründen nicht durchgeführt werden können.

4.3.2. Nicht ausreichende initiale DDA

Bei der Anwendung des miSV besteht die Möglichkeit, daß die initial vorgegebene DDA die tatsächlich zugrunde liegende Krankheit eines Patienten überhaupt nicht erfaßt. Dies zu erkennen kann oft genauso wichtig sein wie die Unterscheidung zwischen den vorgegebenen K-en.¹⁾ Man kann diesen Fall beim miSV berücksichtigen, indem man auf jeder Stufe der DF (Punkt (3) in Diagramm 3.1.) zusätzlich die totale WK $f(x^{(AM)})$ nach Gl. (1-2) bestimmt. Unterschreitet $f(x^{(AM)})$ einen gewissen vorgegebenen Wert, wird das zuletzt berechnete U-Ergebnis gestrichen und das letzte neue Merkmal aus PM nachträglich herausgenommen. Bekommt man im gesamten Verlauf der DF an keiner Stelle eine ausreichende WK für $x^{(AM)}$, wird das eingegebene Ekg mit dem Urteil " Keine der DD-vorgegebenen Krankheiten liegt zugrunde " versehen.

4.3.3. Vorliegen mehrerer Erkrankungen

Ein weiteres Problem entsteht, wenn bei einem Patienten mehrere der DD-vorgegebenen K-en vorliegen. Diesen Fall

1) Von MICHAELIS (1972) wird dies durch die Berücksichtigung von Quantilen bei der Zuteilung einzelner Fälle zu erfassen versucht.

sehen sowohl die eDA als auch - zunächst - die mDA nicht vor, da die beiden Klassifizierungsregeln zugrunde liegenden statistischen Parameter μ_j und Σ_j nur anhand von Ekg-Resultaten solcher Patienten geschätzt wurden, für die genau eine der vorgegebenen K-en zutraf. Will man das Vorliegen mehrerer K-en exakt behandeln, muß man alle 2^N Kombinationen der N K-en vorgeben, was mit einem immens steigenden Aufwand (GREMY, 1977) verbunden ist.

Da dies i.a. nicht realisierbar ist, hat HABBEMA (1976) Bedingungen dafür angegeben, wie man den hier diskutierten Fall auch mit der üblichen initialen DDA (mit N K-en) zumindest teilweise behandeln kann. Dabei zeigt es sich, daß eine mDA, die eine Differenzierung einiger der K-en offen läßt, auch an das gleichzeitige Vorliegen mehrerer dieser K-en denken lassen muß.

4.3.4 Berücksichtigung von Risiken und Kosten einer U

Risiken und Kosten einer U spielen bei dem hier behandelten Beispiel der DD kongenitaler Herzvitien mit Hilfe des Ekgs keine Rolle - sieht man von dem Rechenaufwand ab, der bei der Bestimmung einzelner Ekg-Merkmale etwas größer sein mag. Bei der Benutzung riskanter, invasiver oder sehr teurer U-en aber können diese Faktoren von ausschlaggebender Bedeutung sein. Dann ist für die Auswahl der nächstbesten U im Rahmen einer DF nicht nur die zu erwartende FDR wesentlich, sondern auch das Risiko für den Patienten, das in jedem individuellen Fall sehr unterschiedlich sein kann, und der Preis der U.

Im miSV kann man dies erfassen, indem man bei der U-Auswahl (Punkt (9) in Diagramm 3.1.)

$CRIT_{mi}(q_A, AM, m')$ ersetzt durch

$CRIT_{mi}(q_A, AM, m') + C(m', x^{(AM)})$ (vgl. HESS, 1979A).

Dabei stellt $C(m', x^{(AM)})$ das Risiko bzw. die Kosten dar, das die Durchführung der U m' bei bereits durchgeführten U-en mit Ergebnissen $x^{(AM)}$ für den Patienten in sich trägt.

Bei der DD mit Hilfe von Ekg-U-en ist natürlich

$C(m', x^{(AM)}) = 0$ für alle m' , AM und x anzunehmen.

In anderen Fällen, etwa bei der DD eines Rundherdes in der Lunge (ALPEROVITCH, 1976) wird man die Kosten jeglicher U ebenfalls sehr niedrig veranschlagen und bei einem Verdacht auf Bronchialcarcinom auch eine Thorakotomie häufig nicht scheuen.

In den meisten Situationen spielen die genannten Faktoren aber eine erhebliche Rolle. Dies führt dann dazu, daß man die zur Verfügung stehenden U-en in Risiko- und Kosten-
gruppen aufteilt und diese Gruppen nacheinander abhandelt.

So wird bei der DD des Ikterus die Laparaskopie immer am Ende der DF stehen, meist erweist sich ihre Durchführung dann als überflüssig (KNILL-JONES, 1975).

Für solche riskante U-en muß daher in besonderem Maße gefordert werden, ihren Nutzen im Rahmen eines DD-Problems zuverlässig und anhand statistischer Kriterien zu erfassen.

4.3.5. Sequentielle Auswahl der besten U-Kombinationen

ALPEROVITVCH (1976) hat eine solch sequentielle Auswahl der besten U-Kombinationen erstmals bei der DD von Lebererkrankungen in die CDF eingeführt - allerdings nur unter Berücksichtigung univariater und diskreter statistischer Größen. In das miSV ist ein solches Vorgehen ebenfalls einzubauen, allerdings nur unter erheblicher Vergrößerung des Rechenaufwands. Bei jedem Schritt der DF erforderte die Auswahl der nächstbesten U-Kombination (von L U-en) statt des Vergleichs von $(M-M_a)$ U-en den von $\binom{M-M_a}{L}$ U-Kombinationen (für $M=9$, $M_a=3$, $L=2$ etwa statt 6 Berechnungen 15). Prinzipiell ist jedoch auch diese Variante in das miSV leicht einzufügen.

4.4. Beziehungen des neuen Verfahrens zur ärztlichen

Diagnosefindung

Die Tatsache, daß die hier konstruierten SV-en aus einer Analyse der äDF entwickelt wurden und daß ihre Handhabung eine große Flexibilität und Variabilität aufweist, macht eine Übertragung der Ergebnisse dieser Computer-gestützten Verfahren auf ärztliches Diagnoseverhalten relativ leicht. Dabei entspricht der üblichen äDF wohl am besten das uiSV, wählt doch der Arzt seine U-en i.a. abhängig von der aktuellen DDA (individuell) im Vergleich der jeweils einzeln zu erwartenden Trennfähigkeit (univariat) aus und trifft - bei ausreichender Stabilität der DF - eine DA, die die erzielten Ergebnisse auch in ihren Abhängigkeiten untereinander berücksichtigt (multivariate DA). Es ist gezeigt worden, daß bei einer solchen Anwendung der sequentiellen U-Auswahl eine reservierte Diagnosehaltung in Form der mDA vorteilhaft ist (Kap. 3.3.3.) und daß diese eine zusätzliche Wichtung einzelner K-en meist nicht nötig macht (Kap. 4.2.). Dagegen kann eine Schwerpunktsetzung bei Anwendung einer " Alles-oder-nichts " - Haltung (eDA) vorteilhaft sein. Die Benutzung einer eDA ist allerdings nur solchen Ärzten zu empfehlen, die eine sehr große Anzahl von U-en routinemäßig durchführen, wobei die Auswahl dieser U-en sehr problematisch ist (Kap. 4.1.). Dabei bedeutet ein Mehr an Tests keineswegs zwangsläufig ein Mehr an diagnostischer Sicherheit.

Die hier nur aufgezeigten, jedoch nicht an Beispielen

erläuterten, weiteren Variationsmöglichkeiten der SV-en mit mDA (Kap. 4.3.) können ebenfalls nützliche Hinweise auf ein geeignetes ärztliches Diagnoseverhalten geben. Bei der Einführung dieser Variationen in das miSV bereiten in praxi der Ausschluß gewisser U-en, die Berücksichtigung des Auftretens von initial nicht vorgesehenen K-en und das Erfassen gleichzeitig vorliegender K-en keine Probleme, wohl aber die Quantifizierung von Risiken und Kosten einer U und die sequentielle Auswahl von U-Kombinationen.

Die drei grundsätzlich neuen Gesichtspunkte des hier vorgestellten Computer-gestützten Verfahrens zur DF,

- multivariat individuelle U-Auswahl,
- Berücksichtigung der Stabilität der U-Ergebnisse
- flexibles Diagnoseverhalten in Form der mDA

bei strikter Beibehaltung und Anwendung entscheidungstheoretischer Grundsätze bieten aber nicht nur die Gewähr für äußerst niedrige FDR-en bei minimaler U-Anzahl, sondern eröffnen auch neue Möglichkeiten für eine kombinierte Arzt-Computer-Diagnose.

So bestimmt der Arzt durch die Wahl der initialen DDA, des maximal möglichen Pools an U-en, der Basis-U-en und der Klassifikationskonstanten die Anfangsbedingungen des miSV, auf das er durch die beschriebenen Variationsmöglichkeiten - Wichtung schwerwiegender Erkrankungen, Berücksichtigung von Risiken und Kosten einzelner U-en etc. - weiteren Einfluß nehmen kann. Außerdem hat er die Möglichkeit, auf jeder Stufe der DF durch eine Änderung

des Auswahlkriteriums CRIT seine subjektive Wertung des Diagnoseablaufs in das miSV einzubringen - was allerdings den Optimalitätskriterien möglicherweise entgegensteht.

Außerdem ist es die Aufgabe des Arztes, aus der endgültigen DA die weiteren Schlüsse zu ziehen. Ist diese eindeutig, kann er eine Therapie veranlassen; ist sie aber echt modifiziert und läßt somit eine Differenzierung eines Teils der initial vorgegebenen DDA vermissen, hat er mehrere Möglichkeiten: Stehen ihm für die verbleibende DDA U-en zur Verfügung, die hierfür zwar geeignet erscheinen, für die das miSV aber nicht anwendbar ist, weil die zugrunde liegenden statistischen Parameter nicht zuverlässig geschätzt werden konnten, sollte er diese U-en in konventioneller Weise ausführen. Gibt es auch solche U-en nicht, kann er sich entweder mit der mDA zufrieden geben - möglicherweise hat die verbleibende Unsicherheit keinen Einfluß auf die Therapie - oder er erzwingt eine eindeutige DA durch eine Variation der Klassifikationskonstanten - allerdings auf Kosten eines größeren Fehldiagnoserisikos.

Es soll jedoch betont werden, daß dies in unserem Beispiel nicht nötig ist. Bei der DD kongenitaler Herzvitien kann die nach Ekg-Auswertung (mit Hilfe des miSV mit mDA) verbleibende DDA immer bis zu einer eindeutigen DA fortgeführt werden, etwa durch eine Herzkatheter-U. Wichtig ist allerdings, daß zunächst die leicht verfügbaren U-en die initiale DDA so weit wie möglich eingrenzen, bevor

weitergehende aufwendige oder riskante U-Methoden angewendet werden.

Das hier vorgestellt miSV ist dabei in der Lage, die Differenzierung der DDA unter den Rahmenbedingungen, die der Arzt setzt, mit dem gegebenen U-Pool in optimaler Weise vorzunehmen.¹⁾ So kann der Arzt die CDF an den Platz einsetzen, an dem sie ihm für die gesamte DF am nützlichsten ist. Er kann weiterhin seine Intuition, Ideen und Gefühle einbringen und selbst bestimmen, " to what extent the computer will be able to replace the human brain in medical diagnosis ".²⁾ Diese Tatsache, die mit dem weiteren Vordringen der Computer-Technologie noch an Bedeutung gewinnen wird, zusammen mit der Erkenntnis, daß eine rationelle und sparsame U-Auswahl in Zukunft eine noch größere Rolle spielen wird, läßt das vorgestellte Verfahren noch bedenkenswerter erscheinen.

1) Die Notwendigkeit, vor dem Einsatz Computer-gestützter Verfahren die zugrundeliegende medizinische Fragestellung sorgfältig zu studieren, hat auch ALPEROVITCH (1976) ausdrücklich betont.

2) Vgl. GROSS, 1966

ZUSAMMENFASSUNG

Zu Beginn von Kapitel 1 wird eine kurze Charakterisierung der ärztlichen und der Computer-gestützten Diagnosefindung gegeben, wobei für diese die multivariate Diskriminanzanalyse mit "Bayes"-Entscheidung vorgestellt wird. Dem Vorteil dieser Methode, einer sehr niedrigen Fehldiagnoserate, werden ihre Nachteile, fehlende Variabilität in der Untersuchungsauswahl und in der Form der diagnostischen Aussagen und daher ungenügende Vergleichbarkeit mit dem ärztlichen Diagnosevorgang, gegenübergestellt. Mit einer Computer-gestützten Analyse von fünf verschiedenen Diagnoseverläufen anhand von Beispielen aus der Ekg-Differentialdiagnose kongenitaler Herzvitien werden vier Forderungen an eine optimale Untersuchungsauswahl gestellt: Individuelle Selektion der Untersuchungen, die bei möglichst geringer Untersuchungsanzahl einen gleichmäßigen Verlauf der Diagnosefindung gewährleistet und zu einer ausreichend differenzierten und die zutreffende Krankheit erfassenden diagnostischen Aussage führt.

In Kapitel 2 wird die "modifizierte" diagnostische Aussage eingeführt, die es erlaubt, in gewissen Fällen auf eine Differenzierung zwischen bestimmten Krankheiten zu verzichten. Damit erreicht man bei der Differentialdiagnose kongenitaler Herzvitien eine deutliche Reduktion der Fehldiagnoserate - die der "Bayes"-Entscheidung ("eindeutige" diagnostische Aussage) wird bis auf ihren vierten Teil reduziert. Daneben erleidet man jedoch einen Verlust

an diagnostischer Schärfe, der quantifiziert und in einem dritten Kriterium, dem "Modifikationsquotienten", zu der Reduktion der Fehldiagnoserate in Beziehung gesetzt wird.

In Kapitel 3 wird in die Computer-gestützte Diagnostik die schrittweise Auswahl von Untersuchungen - vorgenommen anhand verschiedener Optimalitätskriterien - eingeführt. Beim Vergleich dieser "sequentiellen" Selektionsverfahren mit den herkömmlichen "nicht-sequentiellen" zeigt sich, daß jene, besonders in ihrer "multivariat individuellen" Form, schon mit zwei Drittel der maximal möglichen Untersuchungen die minimale Fehldiagnoserate ergeben. Dies gilt sowohl für die Anwendung einer "eindeutigen" als auch - und in besonderem Maße - einer "modifizierten" diagnostischen Aussage: Hiermit erreicht man eine minimale Fehldiagnoserate bei weitestgehender Differenzierung der diagnostischen Aussage und maximaler Ersparnis an Untersuchungen.

In Kapitel 4 werden weitere Variationsmöglichkeiten des neuen Verfahrens vorgestellt. Dazu wird zunächst gezeigt, daß unsere Methode weitgehend unabhängig von der Wahl der Basisuntersuchungen ist. Weiterhin wird diskutiert, daß man durch eine Wichtung der differentialdiagnostisch vorgegebenen Krankheiten das Nichterkennen besonders wichtiger Erkrankungen bei Anwendung der "eindeutigen" diagnostischen Aussage verhindern kann, daß aber die "modifizierte" Aussage dieses Ziel ohne Wichtung und unter gleichzeitiger Minimierung der durchschnittlichen Fehldiagnoserate auch erreicht. Kapitel 4.3 zeigt, daß in

unsere Methode auch andere Variationen leicht eingefügt werden können: Ausschluß gewisser Untersuchungen, Erkennen einer nicht ausreichenden initialen differentialdiagnostischen Fragestellung und des gleichzeitigen Vorliegens mehrerer Erkrankungen, Berücksichtigung von Kosten und Risiken der Untersuchungen und die sequentielle Auswahl von Untersuchungskombinationen. In Kapitel 4.4. wird anhand der Computer-gestützten Analyse aufgezeigt, unter welchen Bedingungen auch bei der ärztlichen Diagnosefindung ein modifiziertes Diagnoseverhalten angemessen ist und demonstriert, daß das neue Verfahren eine der ärztlichen Diagnosefindung ähnliche Flexibilität ermöglicht und eine vom Arzt gesteuerte Kooperation Arzt-Computer vielfacher Hinsicht erlaubt.

Liste der Abkürzungen

äDF	: S. ii, 2
AM	: S. 6
ASD	: S. 4
BE	: S. 9
BK	: S. 43
CDF	: S. ii, 2
DA	: S. i, 1
DD	: S. i, 4
DDA	: S. i, 1
DF	: S. i, 1
DS	: S. 23
eDA	: S. iv, 8
FDR	: S. i, 3
FT	: S. 4
K	: S. 21
mDA	: S. v, 13
md	: S. 48
mi	: S. viii, 46
MQ	: S. vi, 32
NSV	: S. iv, 25
PM	: S. 6
PS	: S. 4
SR	: S. 41
SV	: S. iv, 17
U	: S. i, 1
ud	: S. 48
ui	: S. 48
VaDS	: S. v, 22
VSD	: S. 4
WK	: S. 21

Liste der Bezeichnungen

c_j	: S. 21
ΔFDR	: S. 32
ΔINF	: S. 24
δ_B	: S. 9
δ_E	: S. 8
δ_M	: S. 21
FDR_B	: S. 9
FDR_{δ_E}	: S. 9
$FDR_{\delta_E}(j)$	: S. 8
$FDR_{\delta_E}(i j)$	: S. 8
FDR_M	: S. 22
$FDR_M(k)$	: S. 22
$FDR_M(J k)$	: S. 21
$f(x^{(AM)})$	: S. 6
$f(x^{(AM)} j)$	: S. 6
I	: S. 21
J	: S. 21
INF_M	: S. 23
M_a	: S. 6
$\mu_j^{(AM)}$	: S. 7
n_M	: S. 22
p_j	: S. 6
$q(j x^{(AM)})$	: S. 6
Q_w	: S. 32
$\sum_j^{(AM)}$	: S. 7
$TINF_E$	: S. 24
$x^{(AM)}$	: S. 6

LITERATURVERZEICHNIS

- ALPEROVITCH, A., LE MINOR, M., LELLOUCH, J. (1976):
Three examples of Computer-aided Medical Decision.
In DOMBAL, F.T., GREMY, F. (Eds): Decision making
and medical care. Can Information Science help?,
pp. 143-149. North Holland Pub. Comp.
- ALPEROVITCH, A., LE MINOR, M., LELLOUCH, J. (1977):
The choice of a strategy for the selection of comp-
lementary investigations: A study by simulations.
In SHIRES, D.B., WOLF, H. (Eds): MEDINFO 77, pp. 133
-136. Amsterdam: North Holland Publ. Comp.
- BAY, K.S., LEE, S.J.K., FLATHMAN, D.P., ROLL, J.W., PIERCY,
W. (1976): Application of step-wise Discriminant
Analysis and Bayesian Classification Procedure in
determining prognosis of Acute Myocardial Infarction.
Canad. med. Ass. J. 115, pp. 887-892.
- BRODDA, K., HESS, C.F. (1981): Kriterien zur optimalen
Klassenauswahl bei Computer-gestützten Mustererken-
nungsverfahren. Zur Veröffentlichung eingereicht.
- BROOKS, R.C., DWYER, S.J., LODWICK, G.S. (1974): Com-
puter Diagnosis of Congenital Heart Disease. Proc.
of the 1974 International Conf. on Systems, Man, and
Cybernetics, Qct.2-4, Dallas, Texas. pp. 13-20.
- CORNFIELD, J., DUNN, R.A., BATCHLOR, C.D., PIPBERGER, H.V.
(1973): Multigroup Diagnosis of Electrocardiograms.
Comput. Biomed. Res. 6, pp. 97-120.

DOMBAL, F.T., HORROCKS, J.C., WALMSLEY, G., WILSON, P.D.

(1975): Computer-aided Diagnosis and Decision-making
in the Acute Abdomen. J. Roy. Coll. Physicians of
London, Vol.9, pp. 211-218.

GREMY, F., SALAMON, R., CHASTANG, C. (1977): Theory

of Medical Decision Processes. Introductory Remarks.

In SHIRES, D.B., WOLF, H. (Eds): MEDINFO 77, pp. 95
-99.

GROSS, R. (1966): From Intuition to Computation. Developement
and Problems of Medical Diagnosis. Meth. Inf.

Med. 5, pp. 35-39.

HABBEMA, J.D.F. (1976): Models for Diagnosis and Detection

of Combinations of Diseases. In DOMBAL, F.T., GREMY, F.

(Eds): Decision Making and Medical Care. Can In-
formation Science help?, pp. 399-410. North Holland
Publ. Comp.

HESS, C.F. (1979A): Die Klassenauswahl im Mustererken-

nungsprozessen - dargestellt am Beispiel Computer-
gestützter Diagnoseverfahren in der Medizin.

Dissertation, Mainz

HESS, C.F., BRODDA, K. (1979B): On the Optimum Choice

of categories for the classification of biomedical
data patterns. Meth. Inform. Med. 18, 222-228

HESS, C.F., BRODDA, K., WELLNER, U. (1979C): On the Con-

struction of Optimum Categories in Biomedical Data
Recognition Problems. Proc. Medical Informatics 2nd
Eur. Fed. Conf., Berlin, pp. 628-634

- HESS, C.F., BRODDA, K. (1980): A Computer Program
Suitable for Analysis of Choice of Categories in
Biomedical Data Recognition Problems. Comp. Programs
in Biomedicine 12, pp. 55-59
- HESS, C.F., BRODDA, K. (1981): Estimation of the Ex-
pected Error Rate when Transferring a Diagnosis
System from one Hospital to another. Comp. and
Biomed. Research (zur Veröffentlichung eingereicht)
- JACQUES, J.A., NORUSIS, M.J. (1976): The Importance of
Symptom Non-independence in Diagnosis. In DOMBAL, F.T.,
GREMY, F. (Eds): Decision Making and Medical Care.
Can Information Science help?, pp. 379-390. North
Holland Publ. Comp.
- KNILL-JONES, R.P. (1975): The Diagnosis of Jaundice by
the Computation of Probabilities. J. Roy. Coll. Phycns
London 9, pp. 205-210
- KNILL-JONES, R.P., STERN, R.B., GIRMES, D.H. MAXWELL, J.D.,
THOMPSON, R.P., WILLIAMS, R. (1973): Use of
Sequential Bayesian Model in Diagnosis of Jaundice
by Computer. Brit. Med. J. 1973,I,pp. 530-533
- LACHENBRUCH, P.A. (1977): The Effect of Non-normality
on the Quadratic Discrimination Function. In SHIRES,
D.B., WOLF, H. (Eds): MEDINFO 77, pp. 101-104.
Amsterdam: North Holland Publ. Comp.
- VON MENGDEN, H.J. (1977A): Computereinsatz in der medi-
zinischen Diagnostik, dargestellt am Beispiel des
Elektrokardiogramms. Ärzteblatt Rheinl.-Pfalz 12,
pp. 1138- 1145

- VON MENGDEN, H.J. (1977 B): Computer-Diagnostik des Elektokardiogramms. Dtsch. med. Wochenschrift 102, pp. 529-534
- MICHAELIS, J. (1972): Zur Anwendung der Diskriminanzanalyse für die medizinische Diagnostik. Habilitationsschrift, Mainz
- MOON, R.E., JORDANOV, S., PEREZ, A., TURKSEN, I. (1977): Medical Diagnostic System with Human-like Reasoning Capability. In SHIRES, D.B., WOLF, H. (Eds): MEDINFO 77, pp. 115-119. Amsterdam: North Holland Publ. Comp.
- MUCCIARDI, A.N., GOSE, E.E. (1971): A Comparison of Seven Techniques for Choosing Subsets of Pattern Recognition Properties. IEEE Trans. Computers, Vol. C-20, pp. 1023-1031
- PIPBERGER, H.V. (1973): Methods of Diagnostic ECG Classification. In ZYWIETZ, C., SCHNEIDER, B. (Eds): Computer Application on ECG and VCG Analysis. North Holland Publ. Co., pp. 296-306
- PIPBERGER, H.V. (1978): Potentialities, Problems and Perspectives of ECG Data Processing. Verh. Dtsch. Ges. Kreislaufforschg. 44, pp. 33-37
- PÖPPL, S.J., HERRMANN, G., HÖBEL, W., STEINMÜLLER, F. (1977): Adaptive Nonlinear Approaches to Nonparametric Classification and Discrimination Techniques Based on Pipberger ECG-file. In SHIRES, D.B., WOLF, H. (Eds): MEDINFO 77, pp. 1017-1021. Amsterdam: North Holland Publ. Comp.

- QUESENBERRY, C.P., GESSAMAN, M.P. (1968): Nonparametric Discrimination Using Tolerance Regions. Ann. Math. Stat., Vol. 39, No.2, pp. 664-673
- SMETS, P., VAINSEL, H., BERNARD, R., KORNREICH, F. (1977): Bayesian Probability of Fuzzy Diagnosis. In SHIRES, D.B., WOLF, H. (Eds): MEDINFO 77, 121-122 Amsterdam: North Holland Publ. Co.
- STERN, R.B., KNILL-JONES, R.P., WILLIAMS, R. (1974): Clinician versus Computer in the Choice of 11 Differential Diagnoses of Jaundice Based on Formalized Data. Meth. Inform. Med. 13, pp. 79-82.
- WATANABE, Y., NAKAMURA, A., TANIMURA, H., YOKOI, M., SOBOBATA, I., YASUI, S. (1976): Computer Diagnosis of Congenital Heart Disease by Means of Vector Cardiograms Using the Likelihood Method. In DE DOMBAL, F.T., GREMY, F. (Eds): Decision Making and Medical Care. Can Information Science Help?, pp. 185-193. Amsterdam: North Holland Publ. Co.

LEBENS LAUF

Am 9.3.1951 wurde ich, Clemens Friedrich Heß, als drittes von fünf Kindern des Arztes Dr. Anton Heß und seiner Ehefrau, der Ärztin Margarete Heß, geb. Loenartz, in Koblenz geboren. Nach dem Besuch der Volksschule und des neusprachlichen Gymnasiums in Cochem bestand ich dort 1969 das Abitur.

Im WS 1969/70 begann ich an der Universität Mainz das Studium der Mathematik mit Nebenfach Physik, legte im WS 1971/72 die Diplom-Vorprüfung, im WS 1975/76 die Diplomhauptprüfung ab und promovierte 1979 am Fachbereich Mathematik der Universität Mainz zum Doktor der Naturwissenschaft. Im SS 1975 begann ich an der Universität Mainz das Studium der Humanmedizin, bestand 1977 die ärztliche Vorprüfung, 1978 den ersten, 1980 den zweiten und 1981 den dritten Teil der ärztlichen Prüfung. Im Mai 1981 erhielt ich die Approbation zum Arzt.

Ich bin verheiratet mit der Studentin der Romanistik, Germanistik und Musikwissenschaft, cand. phil. Sabine Heß, geb. Bungart und habe ein Kind. Seit Juli 1981 arbeite ich als Assistenzarzt in der Klinik für Nuklearmedizin und Radiotherapie des Universitätsspitals Zürich.

Zürich, den 27.10.81

D. GRÄBNER

DISSERTATIONS U. OFFSETDRUCKEREI



Gotenstraße 3
8602 Altendorf bei Bamberg
Telefon. 09545 / 8791